(* COUNTS *)

# 28

## Expressing Counted Fractions


Introduction 28-1
- relative importance 28-1

28A. Three matched scales for counted fractions 28-2
- half "cigars" 28-9

28B. An opinion example where careful expression clearly pays off 28-9a

28C. Diagnosing needs for changed expression 28-14
- shifts 28-23
- finding the shift 28-25

28D. Mountains and mice (back to Nebraska) 28-29
- behavior of a composite 28-38

28E. Straightening response curves 28-41
- delicacy and vigor in re-expression 28-44

28F. Double dichotomies 28-48

28G. Counts of counted fractions 28-50

**********

28H. How far have we come? 28-60

**********

28K. Easy f-roots and f-logs with a slide rule 28-62

**********

28L. Comparison of simple and refined expressions for counted fractions (background) 28-65

**********

28M. Fitting comparison curves related to folded powers (optional) 28-71

**********

28P. Problems for chapter 28 28-79

**********

Tables for reference

Matched pluralities, f-roots and f-logs 28-5/5a
Widths: split "cigars" 28-11
Shifts: to f-roots 28-26
to f-logs 28-27/28'
Fitting λ-distribution comparison curves 28-77/78


| Exh | Page |
|---|---|
| a | |
| 1 | 28-5 |
| 1a | 28-5a |
| 2 | 28-7 |
| 3 | 28-8 |
| b | |
| 4 | 28-10 |
| 5 | 28-11 |
| c | |
| 6 | 28-15 |
| 7 | 28-16 |
| 8 | 28-18/19 |
| 9 | 28-20/21 |
| 10 | 28-22 |
| 11 | 28-24 |
| 12 | 28-26 |
| 13 | 28-27 |
| 14 | 28-28 |
| d | |
| 15 | 28-30/32 |
| 16 | 28-34/36 |
| 17 | 28-39/40 |
| e | |
| 18 | 28-42 |
| 19 | 28-45 |
| 20 | 28-46 |
| f | |
| 21 | 28-49 |
| g | |
| 22 | 28-51 |
| 23 | 28-52/55 |
| 24 | 28-56 |
| 25 | 28-57 |
| 26 | 28-59 |
| h | |
| k | |
| l | |
| 27 | 28-68 |
| 28 | 28-70 |
| m | |
| 29 | 28-73 |
| 30 | 28-74 |
| 31 | 28-76 |
| 32 | 28-77/78 |
| p | |
| 33 | 28-80 |

3Ie 28

# Expressing Counted Fractions

A count comes to a forced halt in one place, zero. A counted fraction comes to a forced halt at two places

zero and one

The expressions that we use for counted fractions need to take both these forced halts into account, as well as the symmetry between a fraction and its complement -- for example, between 13% and 87%.

In dealing with counted fractions, then, we can expect to see all the sorts of things go on that go on for counts, only on both ends, symmetrically. The details will, of course, be somewhat different.

## * relative importance *

What can we say about the relative importance of changes in counted fractions? Their usual expression is in %. We often say that 20% of all accidents have a certain type of cause. Or that 95% of a certain population can read a newspaper. Consider trying to change some simple aspect of people: can they read a newspaper, do they wear beards, do they buy "Stistix" breakfast food. If 41% can read now, a campaign to raise this figure by 1% to 42% (or in many cases to lower it by 1% to 40%) is likely to be no great task. But if only 1% -- or as many as 97% -- can

read, to raise 1% to 2% -- or 97% to 98% -- is likely to require much effort.

Except in determining the outcome of elections, a 1% change is a much greater thing near 0% or 100% than it is near 50%. If we are to express counted fractions in a way that reflects this, we must

stretch both ends

of our scale in comparison with the middle. How is it wise to do this?

28A. Three matched scales for counted fractions

First, an important remark: Treating "those who are" and "those who aren't" symmetrically is often very useful -- and sometimes important. This means two things:

- choosing scales with a value of 0 at 50%,
- arranging matters so that a swap of f with 1-f, where f is the fraction concerned, will change the sign, but not the size, of our expression of f.

While we are at it, since $f = \frac{1}{2}$ is to be scored as 0, we may as well plan to have a variety of choices that agree, not only exactly for $f = \frac{1}{2}$ but quite closely for values around $f = \frac{1}{2}$. This is a matter of convenience. It also allows ever-repeated emphasis on how similar all the expressions are "in the middle".

Matching near 50% does not cost us anything, because whether we analyze a given set of numbers or a set

of numbers each exactly three times as large as the numbers of the given set is almost certain to make no essential difference -- when complete, the results can, in any case, be translated to a common, generally understandable scale.

With these guideposts, what can we do that will be simple and effective? The simplest measure that makes $f = \frac{1}{2}$ score 0 without stretching the scale is the "plurality"

$$(f) - (1-f)$$

which is the difference between the fraction of those that are and the fraction of those that aren't.

Another reasonable name for this expression would be folded fraction. Clearly, if we fold any other expression of a fraction -- by subtracting the given function of those who aren't from the same function of those who are -- we will force $f = \frac{1}{2}$ to have score zero and will make the score of 63% the negative of the score of 37%, whatever expression we chose to begin with. Indeed, if we are to have the sort of symmetry around 50% we seek, we can confine our attention to folded expressions. (Proof?)

While a number of more complicated expressions have been -- and may well continue to be -- popular, we can usually do quite well by folding some very simple expression. By analogy with all our other inquiries into good ways to

express data, we find it natural to begin with a square root and a logarithm. We can use the root -- or the log -- of any multiple of f. If we are to match folded fractions near 50%, it turns out that we should fold $\sqrt{2f}$ and $\frac{1}{2}\log_e f \doteq 1.15 \log_{10} f$ (here $\log_e$ indicates the so-called natural or Napierian logarithm -- the logarithm to the base e = 2.71828...).

We need names for these expressions -- both formal names and working names. Clearly <u>folded roots</u> and <u>folded logs</u> are both formal and descriptive, while <u>f-roots</u> (perhaps pronounced "froots") and <u>f-logs</u> (perhaps pronounced "flogs") are convenient. Specifically, then, we plan to consider

$$\text{f-root} = \sqrt{2f} - \sqrt{2(1-f)}$$

$$\begin{aligned}\text{f-log} &= \tfrac{1}{2}\log_e f - \tfrac{1}{2}\log_e(1-f) \\ &= 1.15 \log_{10} f - 1.15 \log_{10}(1-f) \\ &= 1.15 \log_{10}(f/(1-f))\end{aligned}$$

as matched alternatives to

$$\text{plurality} = f - (1-f) .$$

We lay the questions of free starts for logs and modified forms for roots aside for the moment.

Exhibit 1 sets out two-decimal values of matched pluralities, f-roots, and f-logs for every % from 10 to 90 and at

exhibit 1 of chapter 28 -- expressions tabled
Pluralities, folded roots, folded logarithms --
alternative expressions for counted fractions
(Take sign of answer from head of column giving %)

| + | Plur. | f-root | f-log | - | + | Plur. | f-root | f-log | - |
|---|---|---|---|---|---|---|---|---|---|
| 50% | use | .00 | use | 50% | 85% | .70 | .77 | .86 | 15% |
| 51 | → | .02 | ← | 49 | 86 | .72 | .79 | .91 | 14 * |
| 52 | | .04 | | 48 | 87 | .74 | .81 | .95 | 13 |
| 53 | | .06 | | 47 | 88 | .76 | .84 | 1.00 | 12 |
| 54 | | .08 | | 46 | 89 | .78 | .87 | 1.05 | 11 |
| 55% | use | .10 | use | 45% | 90.0% | .80 | .89 | 1.10 | 10.0% |
| 56 | → | .12 | ← | 44 | 90.5 | .81 | .91 | 1.13 | 9.5 * |
| 57 | | .14 | | 43 | 91 | .82 | .92 | 1.16 | 9 |
| 58 | | .1 | | 42 | 91.5 | .83 | .94 | 1.19 | 8.5 |
| 59 | | .18 | | 41 | 92 | .84 | .96 | 1.22 | 8 |
| 60% | use | .20 | use | 40% | 92.5% | .85 | .97 | 1.26 | 7.5% |
| 61 | → | .22 | ← | 39 | 93 | .86 | .99 | 1.29 | 7 * |
| 62 | | .24 | | 38 | 93.5 | .87 | 1.01 | 1.33 | 6.5 |
| 63 | .26 | .26 | .27 | 37 | 94 | .88 | 1.02 | 1.37 | 6 |
| 64 | .28 | .28 | .29 | 36 | 94.5 | .89 | 1.04 | 1.42 | 5.5 |
| 65% | .30 | .30 | .31 | 35% | 95.0% | .90 | 1.06 | 1.47 | 5.0% |
| 66 | .32 | .32 | .33 | 34 | 95.5 | .91 | 1.08 | 1.53 | 4.5 |
| 67 | .34 | .35 | .35 | 33 | 96 | .92 | 1.10 | 1.59 | 4 |
| 68 | .36 | .37 | .38 | 32 | 96.5 | .93 | 1.12 | 1.65 | 3.5 |
| 69 | .38 | .39 | .40 | 31 | 97 | .94 | 1.15 | 1.74 | 3 |
| 70% | .40 | .41 | .42 | 30% | 97.2% | .94 | 1.16 | 1.77 | 2.8% |
| 71 | .42 | .43 | .45 | 29 | 97.4 | .95 | 1.17 | 1.81 | 2.6 |
| 72 | .44 | .45 | .47 | 28 | 97.6 | .95 | 1.18 | 1.85 | 2.4 |
| 73 | .46 | .47 | .50 | 27 | 97.8 | .96 | 1.19 | 1.90 | 2.2 |
| 74 | .48 | .50 | .52 | 26 | 98.0 | .96 | 1.20 | 1.95 | 2.0 |
| 75% | .50 | .52 | .55 | 25% | 98.2% | .96 | 1.21 | 2.00 | 1.8% |
| 76 | .52 | .54 | .58 | 24 | 98.4 | .97 | 1.22 | 2.06 | 1.6 |
| 77 | .54 | .56 | .60 | 23 | 98.6 | .97 | 1.24 | 2.13 | 1.4 |
| 78 | .56 | .59 | .63 | 22 | 98.8 | .98 | 1.25 | 2.21 | 1.2 |
| 79 | .58 | .61 | .66 | 21 | 99.0 | .98 | 1.27 | 2.30 | 1.0 |
| 80% | .60 | .63 | .69 | 20% | 99.2% | .98 | 1.28 | 2.41* | 0.8% |
| 81 | .62 | .66 | .72 | 19 | 99.4 | .99 | 1.30 | 2.55* | 0.6 |
| 82 | .64 | .68 | .76 | 18 | 99.6 | .99 | 1.32 | 2.76* | 0.4 |
| 83 | .66 | .71 | .79 | 17 | 99.8 | 1.00 | 1.35 | 3.11* | 0.2 |
| 84 | .68 | .73 | .83 | 16 | 100.0% | 1.00 | 1.41 | * | 0.0 |

*For more detail beyond1% or 99%, see exhibit 1a, following.

exhibit 1a of chapter 28 -- expressions tabled

Supplementary table for f-logs of fractions
Beyond 1% or 99%

| + | f-log | - | + | f-log | - |
|---|---|---|---|---|---|
| 99.0 | 2.30 | 1.0 | 99.80 | 3.11 | .20 |
| .1 | 2.35 | .9 | .82 | 3.16 | .18 |
| .2 | 2.41 | .8 | .84 | 3.22 | .16 |
| .3 | 2.48 | .7 | .86 | 3.28 | .14 |
| .4 | 2.56 | .6 | .88 | 3.36 | .12 |
| 99.50 | 2.65 | .50 | 99.90 | 3.45 | .10 |
| .52 | 2.67 | .48 | .91 | 3.51 | .09 |
| .54 | 2.69 | .46 | .92 | 3.57 | .08 |
| .56 | 2.71 | .44 | .93 | 3.63 | .07 |
| .58 | 2.73 | .42 | .94 | 3.71 | .06 |
| 99.60 | 2.76 | .40 | 99.95 | 3.80 | .05 |
| .62 | 2.78 | .38 | .96 | 3.91 | .04 |
| .64 | 2.81 | .36 | .97 | 4.06 | .03 |
| .66 | 2.84 | .34 | .98 | 4.26 | .02 |
| .68 | 2.87 | .32 | .99 | 4.61 | .01 |
| 99.70 | 2.90 | .30 | | | |
| .72 | 2.94 | .28 | | | |
| .74 | 2.97 | .26 | | | |
| .76 | 3.01 | .24 | | | |
| .78 | 3.06 | .22 | | | |

Examples:

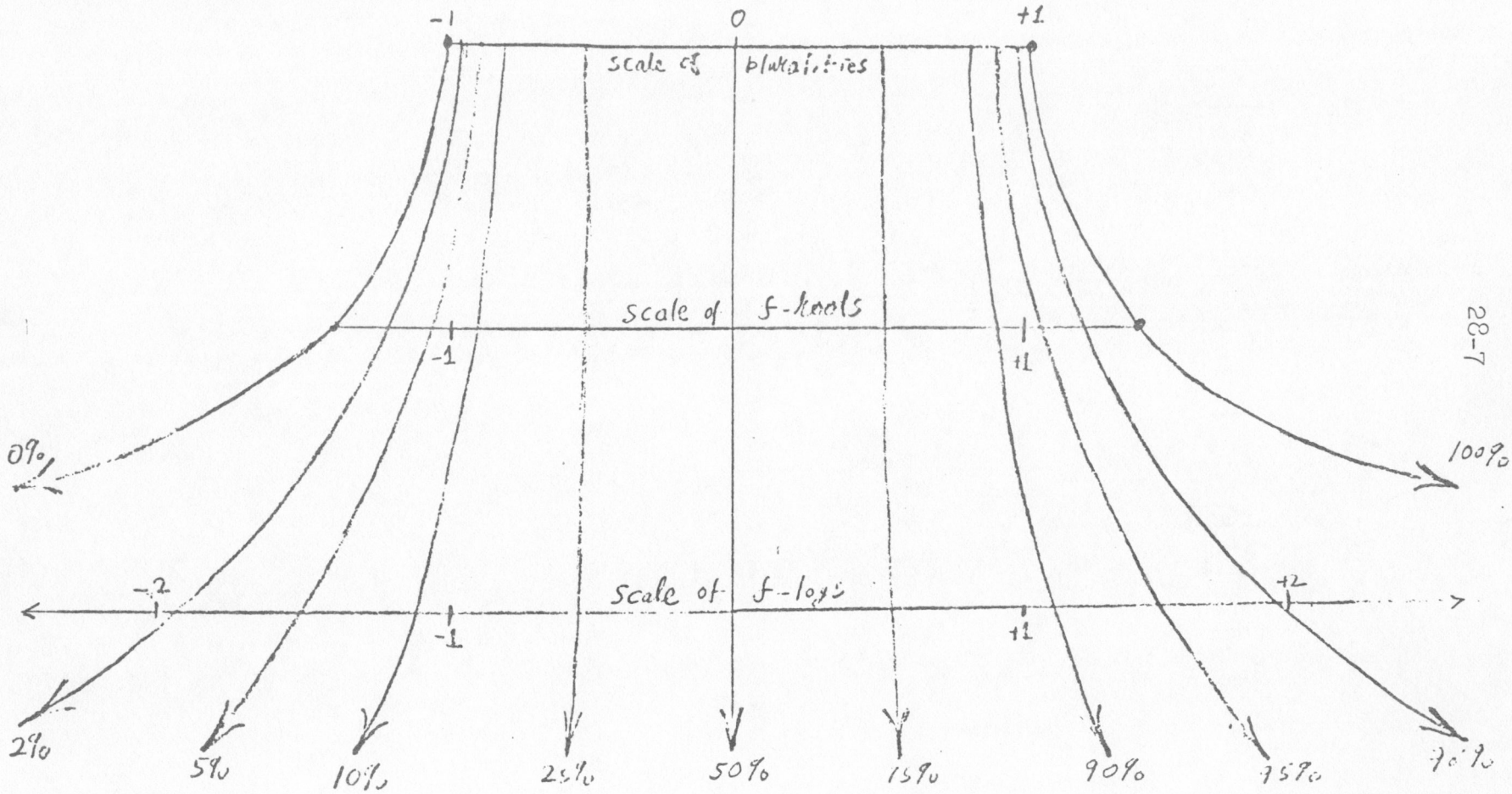
99.29% gives 2.47
0.37% gives -2.80

closer intervals for smaller and larger %'s. Notice that -- to the accuracy given, which is usually sufficient -- the three expressions agree exactly for fractions between 38% and 62%. Unless we have to deal with at least some fractions below 20% or above 80% it is likely to make only trivial differences which of these expressions we use for our counted fractions (so long as we are consistent).

f-roots and f-logs were brought in to stretch the ends of the scale. If the ends of the scale are not used, it is most convenient and satisfactory that f-roots and f-logs behave very, very much like pluralities.

As always it is nice to look at changes of expression graphically as well as in numbers. In this situation, where we think and write of "stretching the tails", it is unusually desirable to compare the different scales side by side and both watch and feel -- to the extent that the feeling can be given -- the tails stretch more and more. Exhibit 2 is intended to do just this.

There is nothing sacred about pluralities. We can match an appropriate version of f-roots or f-logs to anything we like. Exhibit 3 shows the results of matching to plain old percent. It goes so far as to express the matched versions in "%", worrying not about the appearance of both negative % and % greater than 100. These matched expressions are harder to explain -- expecially to the




Picture of the ends being stretched while the center is left undisturbed

exhibit 3 of chapter 25 -- expressions tabled

Another set of matched expressions

| fractions in % | f-roots in % | f-logs in % |
|---|---|---|
| 0 % | -20% | -∞ |
| 0 | -19 | -123 |
| 0.2 | -18 | -106 |
| 0.5 | -16 | -83 |
| 10 | -14 | -65 |
| 2 % | -10% | -48% |
| 5 | -3 | -24 |
| 10 | 6 | -5 |
| 15 | 12 | 7 |
| 20 | 18 | 16 |
| 25 % | 24% | 23% |
| 30 | 30 | 29 |
| 35 | 35 | 34 |
| 40 | 40 | 40 |
| 45 | 45 | 45 |
| 50 % | 50% | 50% |
| 55 % | 55% | 55% |
| 60 | 60 | 60 |
| 65 | 65 | 66 |
| 70 | 70 | 71 |
| 75 | 76 | 77 |
| 80% | 82% | 84% |
| 85 | 88 | 93 |
| 90 | 94 | 105 |
| 95 | 103 | 124 |
| 98 | 110 | 148 |
| 99 % | 114% | 165% |
| 99.5 | 116 | 183 |
| 99.8 | 118 | 206 |
| 99.9 | 119 | 223 |
| 100 % | 120% | ∞ |

uninitiated -- than those of exhibit 1, but they could be just as well used for analysis. (We shall return to them in due course.)

It will often be convenient to use a form of f-log that is not matched to anything, namely

log f - log 1 - f

which equals

log (one count) - log (other count)

This is clearly specially convenient when we start from counts rather than fractions. We call these row f-logs .

* half "cigars" *

If we are going to take f-roots and f-logs seriously, we need to be able to look at them. We want to preserve the area principle, and still have displacement show the f-root or f-log. Exhibit 4 shows how this can be done.

For most purposes the split figures -- on the left in exhibit 4 -- are the more useful since we can show changes clearly by putting a split figure close beside its mirror image.

To make the use of such figures easy, we need a table of relative widths. Exhibit 5 has enough information for sketching in such half "cigars". Scale locations for 75% and 90% are shown in this table, too, because they -- together with 50%, 25% and 10% -- are often the convenient markings to place on a half-cigar.

(The split figure for f-roots is quite cigar like, but that for f-logs would be a very unusual cigar. On the whole, "cigar" seems to be as suggestive a name as has been thought of.)

28B. <u>An opinion example where careful expression clearly pays off</u>

Being careful in how one expresses counts or counted fractions gives us better analyses in many, many instances. It is not always easy to be sure that this is happening to us, often the improvement is not obvious. If the data contains no two portions whose analysis we can compare, it is hard to be sure that we have indeed gained by care in choosing the expression to be analyzed. Even if one portion can be checked against another in which everything is very nearly the same, we are likely to learn little. In such a situation, the effect of a choice of expression on the two portions will be almost the same, consequently the comparison of the two for any one expression is likely to be very similar to that for any other.

exhibit 4 of chapter 28 -- standard displays

Preserving (relative) area while re-expressing vertical displacement

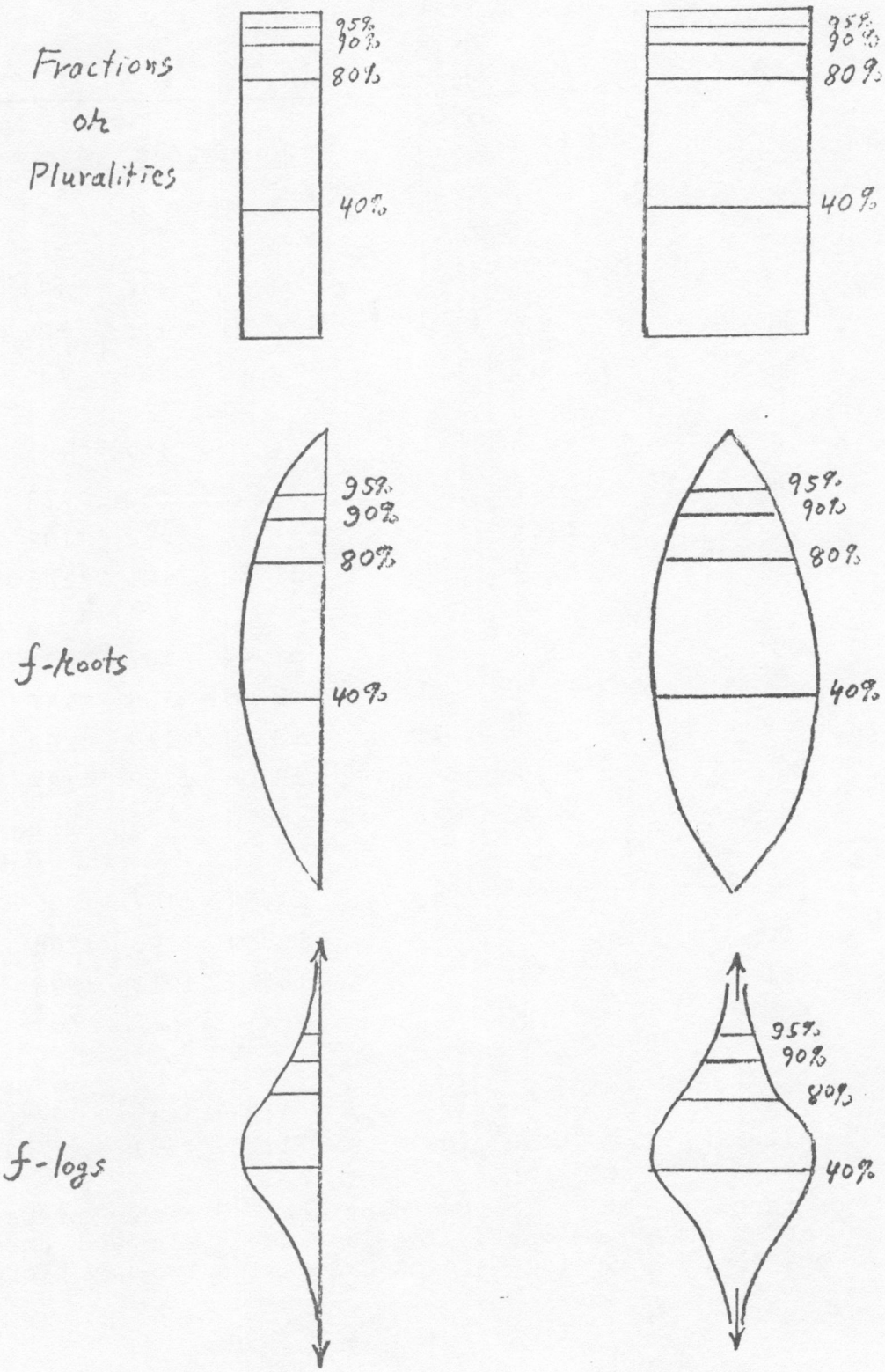


exhibit 5 of chapter 28 -- display widths

Relative widths at various heights that keep areas proportional to counts

| Relative width | Relative height: For f-roots | | Relative height: For f-logs | | |
|---|---|---|---|---|---|
| 1.00 | ±.00 | 0 | ±.00 | ±0 | ±0 |
| .95 | ±.36 | ±25 | ±.23 | ±12 | ±16 |
| .90 | ±.50 | ±35 | ±.33 | ±16 | ±23 |
| .85 | ±.62 | ±44 | ±.41 | ±20 | ±29 |
| .80 | ±.71 | ±50 | ±.48 | ±24 | ±34 |
| .6 | ±.98 | ±69 | ±.75 | ±37 | ±53 |
| .4 | ±1.16 | ±82 | ±1.03 | ±52 | ±73 |
| .2 | ±1.31 | ±93 | ±1.44 | ±72 | ±102 |
| .1 | ±1.36 | ±96 | ±1.82 | ±91 | ±128 |
| | | | | | ± |
| .08 | ±1.37 | ±97 | ±1.94 | ±97 | ±137 |
| .06 | ±1.38 | ±98 | ±2.08 | ±104 | ±147 |
| .04 | ±1.39 | ±98h | ±2.29 | ±114 | ±162 |
| .02 | ±1.40 | ±99 | ±2.64 | ±132 | ±187 |
| .01 | ±1.41 | ±99h | ±2.99 | ±149 | ±211 |
| .005 | " | " | ±3.34 | ±167 | ±236 |
| .002 | " | " | ±3.80 | ±190 | ±268 |
| .001 | " | " | ±4.20 | ±210 | ±296 |
| .000 | ±1.41 | ±100 | (∞) | (∞) | (∞) |
| (75% falls at) | (.52) | (36h) | (.55) | (27h) | (39) |
| (90% falls at) | (.89h) | (63) | (1.10) | (55) | (77h) |

Note: Two or three height scales are provided for each expression as a matter of convenience. The last scale, in each case, will be specially useful in connection with two-way plots.

To learn from internal comparisons -- from comparisons within one set of data, we need a situation where we can compare the behavior of one portion with that of another where:

- our attention is directed toward differences so that satisfactory agreement means similar differences.
- the numbers whose differences are formed -- and found to agree -- are smaller in one portion, larger in the other.

The simplest such situation, of course, is a two-way table in which both factors show substantial effects.

Thinking about the choice of form of expression in section 24A -- thinking about wolves, or a substitute -- led us to consider what small changes were "similar". This is essentially a two-way table approach, although we did not say so -- and likely did not think so. (In the wolf example, for instance, the two-way table is a 2×2 with all combinations of two years and two counties.) Looking at data to see how much careful choice of expression helps -- looking at hopefully comparable differences -- also leads us to a two-way analysis approach. The similarity is far from accidental.

Let us start with a simple 2x2 example. Under the headline "Protestants shift support to Kennedy" the 2 February '62 Washington Post reported the following Gallup poll results

| | Protestants | | Catholics | |
|---|---|---|---|---|
| | Nov. '60 | Jan. '62 | Nov. '60 | Jan. '62 |
| Kennedy | 38% | 59% | 78% | 89% |
| Nixon | 62% | 41% | 22% | 11% |

From an engineering (i.e., practical political) viewpoint, George Gallup was correct in saying that "one of the major -- reasons for the President's popularity has been his success in allaying anti-Catholic sentiment while not losing the support of his fellow Catholics."

But we might want to try to understand (numerically) what is really going on. We can be very sure that this situation demands scale stretching near the ends. (It is surely easier to change 38% to 42% than 89% to 93%.) So let us go to f-logs, and look at the same data again (with f = fraction for Kennedy). The f-logs are:

| | Protestants | Catholics | Difference |
|---|---|---|---|
| Nov. '60 | -.24 | +.63 | +.87 |
| Jan. '62 | +.18 | +1.05 | +.87 |
| Change | up .42 | up .42 | |

so that the Protestant and Catholic changes are very similar and mutually supporting.

Understanding of what happened to John F. Kennedy's popularity between November 1960 and January 1962 -- a scientific matter -- is much better served by the f-log analysis. We can say: "JFK's popularity improved by +0.42 on the f-log scale; you would have got this same answer had you asked only Catholics or only Protestants.

It is this constancy present in f-logs, but not in % or pluralities -- that shows forth the advantages in this example of our choice of f-logs.

We need to see this result -- not merely see its numbers. Exhibit 6 shows it to us in split figures.

One virtue of this type of plot is that it can show both the scientific and the engineering result. This is perhaps clearest in exhibit 7, where we have interchanged roles -- putting one religion on each split figure. We now see both the equal change in f-roots, shown by equal vertical motion, and the unequal change in %, shown by the unequal areas. We could make this picture even more politically practical if we used different widths for the two split figures, choosing widths in proportion to numbers of voters of the two religions.

28C. <u>Diagnosing needs for changed expression</u>

When the response in a two-way table is an amount or a count, we have found it easy to diagnose the need for

exhibit 6 of chapter 28 -- Kennedy support

Change in support for John F. Kennedy

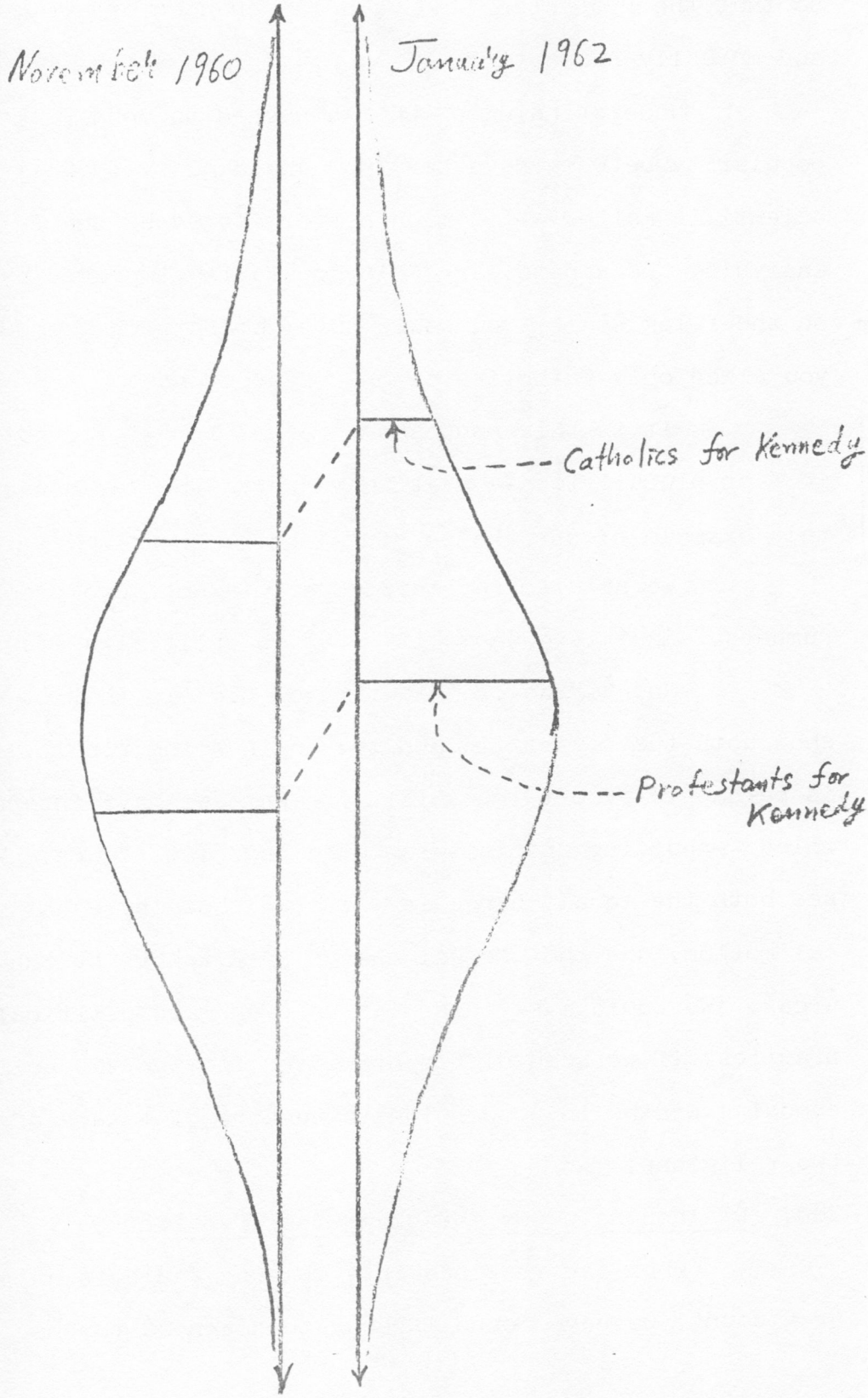
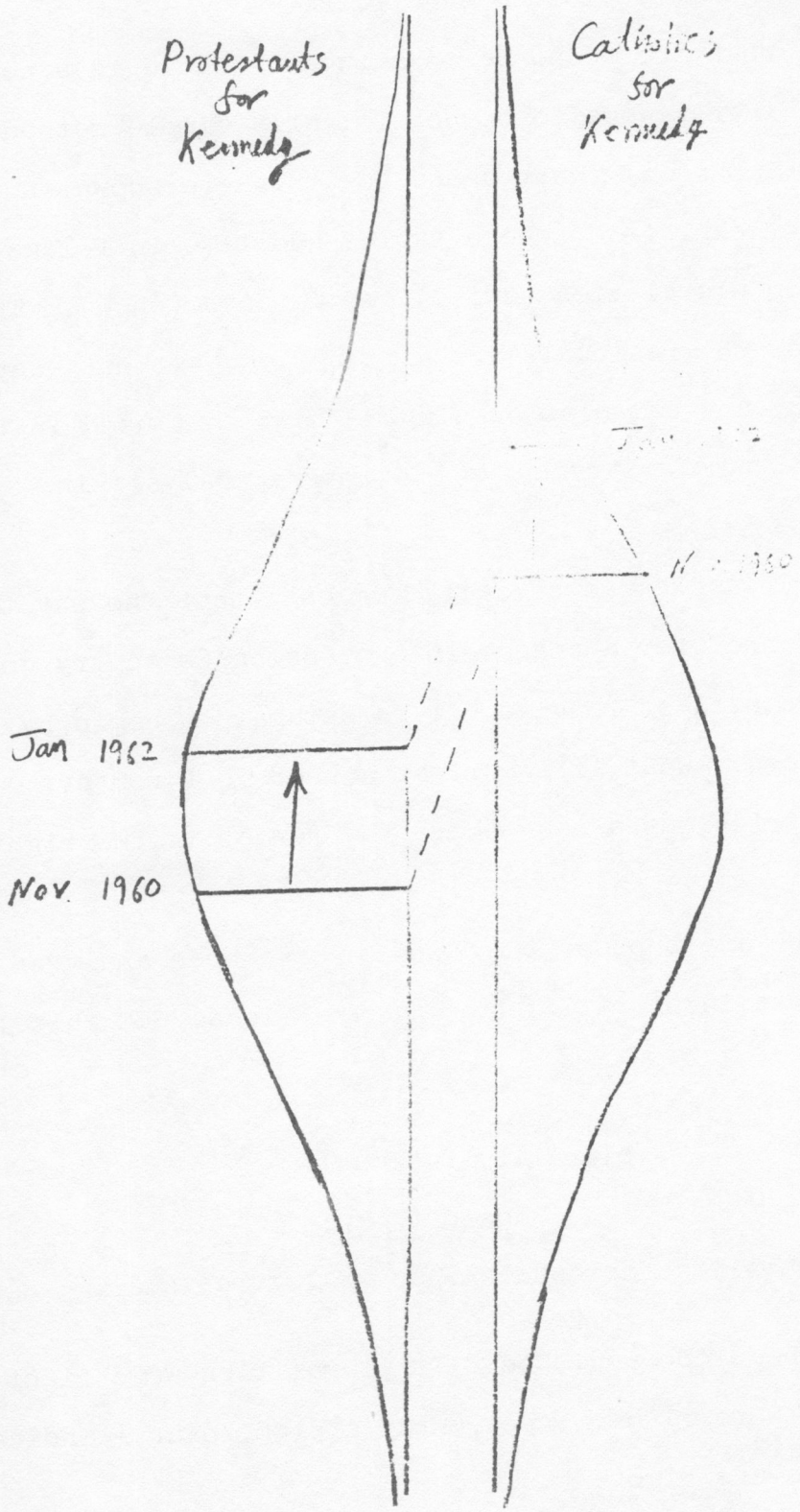


exhibit 7 of chapter 28 -- Kennedy support

Religious difference in support for John F. Kennedy



one of the common changes of expression. Comparison values and the diagnostic plot do the trick.

When the response in a two-way table is a counted fraction, there seems to be no similarly simple method of diagnosis. Some of the reasons for this are shown in the examples of exhibit 8, where tables additive in f-logs are analyzed in pluralities (twice) and f-roots (once). The patterns of the residuals are distinctive, but not easy to forecast by any calculation simpler than that used in the exhibit -- namely, trying an elementary analysis in re-expressed terms.

This example strongly suggests that the way to diagnose a need for a change in expression is to try one. For distinctness of results, it is natural to begin by trying f-logs. What if the right answer is f-roots? Exhibit 9 tries out a case where the f-root is the right answer.

We may suppose the data to have been analyzed twice -- once in pluralities, once in f-logs. Exhibit 10 plots

plurality residual

against

plurality residual minus f-log residual.

The slope would be 0 if pluralities were right and 1 if f-logs were right. Apparently a slope near $\frac{1}{4}$ indicates that f-roots are right.

exhibit 8 of chapter 28 -- casual examples

How need for re-expression can show itself

A) DATA in f-LOGS -- neatly additive, of course -- unit 0.01

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| a | -120 | -90 | -60 | -30 | 0 |
| b | -90 | -60 | -30 | 0 | 30 |
| c | -60 | -30 | 0 | 30 | 60 |
| d | -30 | 0 | 30 | 60 | 90 |
| e | 0 | 30 | 60 | 90 | 120 |

B) DATA in PLURALITIES -- unit 1%

| | | | | | |
|---|---|---|---|---|---|
| a | -83 | -72 | -59 | -29 | 0 |
| b | -72 | -54 | -29 | 0 | 29 |
| c | -54 | -29 | 0 | 29 | 54 |
| d | -29 | 0 | 29 | 54 | 72 |
| e | 0 | 29 | 54 | 72 | 89 |

C) ELEMENTARY ANALYSIS of B -- by means

| | | | | | | eff | fit |
|---|---|---|---|---|---|---|---|
| a | 13 | 1 | -6 | -6 | 0 | -48 | -48 |
| b | 1 | -4 | -4 | 0 | 6 | -25 | -25 |
| c | -6 | -4 | 0 | 4 | 6 | 0 | 0 |
| d | -6 | 0 | 4 | 4 | -1 | 25 | 25 |
| e | 0 | 6 | 6 | -1 | -13 | 48 | 48 |
| eff | -48 | -25 | 0 | 25 | 48 | 0 | 00 |
| fit | -48 | -25 | 0 | 25 | 48 | 00 | 0 |

D) SAME DATA in f-ROOTS -- unit 0.01

| | | | | | |
|---|---|---|---|---|---|
| a | -95 | -79 | -56 | -29 | 0 |
| b | -79 | -56 | -29 | 0 | 29 |
| c | -56 | -29 | 0 | 29 | 56 |
| d | -29 | 0 | 29 | 56 | 79 |
| e | 0 | 29 | 56 | 79 | 95 |

exhibit 8 (Cont'd.) - 2

E) ELEMENTARY ANALYSIS of panel D

| | 1 | 2 | 3 | 4 | 5 | eff | fit |
|---|---|---|---|---|---|---|---|
| a | 9 | 0 | -4 | -4 | 0 | -52 | -52 |
| b | 0 | -2 | -2 | 0 | 4 | -27 | -27 |
| c | -4 | -2 | 0 | 2 | 4 | 0 | 0 |
| d | -4 | 0 | 2 | 2 | 0 | 27 | 27 |
| e | 0 | 4 | 4 | 0 | -9 | 52 | 52 |
| eff | -52 | -27 | 0 | 27 | 52 | 0 | 00 |
| fit | -52 | -27 | 0 | 27 | 52 | 00 | 0 |

F) ANOTHER set of DATA in f-LOGS -- still additive -- unit 0.01

| | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|
| f | -240 | -210 | -180 | -150 | -120 |
| g | -190 | -160 | -130 | -100 | -70 |
| h | -140 | -110 | -80 | -50 | -20 |
| i | -90 | -60 | -30 | 0 | 30 |
| j | -40 | -10 | 20 | 50 | 80 |

G) The DATA of panel F in PLURALITIES -- unit 1%

| | | | | | |
|---|---|---|---|---|---|
| f | -98 | -97 | -95 | -90 | -83 |
| g | -96 | -92 | -86 | -76 | -61 |
| h | -89 | -80 | -73 | -46 | -20 |
| i | -72 | -54 | -29 | 0 | 29 |
| j | -38 | -10 | 20 | 46 | 66 |

H) ELEMENTARY ANALYSIS FOR panel G -- by means

| | | | | | | eff | fit |
|---|---|---|---|---|---|---|---|
| f | 25 | 13 | 1 | -11 | -26 | -44 | -93 |
| g | 16 | 9 | 1 | -14 | -13 | -35 | -84 |
| h | 3 | -1 | -8 | 2 | 6 | -13 | -62 |
| i | -17 | -12 | -1 | 11 | 18 | 24 | -25 |
| j | -25 | -10 | 6 | 15 | 13 | 66 | 17 |
| eff | -30 | -17 | -3 | 14 | 36 | -49 | 00 |
| fit | -79 | -66 | -52 | -35 | -13 | 00 | 49 |

exhibit 9 of chapter 28 -- casual example

When data additive in f-roots is
analyzed in pluralities and f-logs

A) The DATA in f-ROOTS -- my, how additive -- unit 0.01

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| a | -130 | -102 | -74 | -46 |
| b | -100 | -72 | -44 | -16 |
| c | -70 | -42 | -14 | 14 |
| d | -40 | -12 | 16 | 44 |
| e | -10 | 18 | 46 | 74 |
| f | 20 | 48 | 76 | 104 |

B) The DATA of panel A in PLURALITIES -- unit 1%

| | | | | |
|---|---|---|---|---|
| a | -99 | -88 | -69 | -45 |
| b | -86 | -67 | -43 | -16 |
| c | -66 | -41 | -14 | 14 |
| d | -39 | -12 | 16 | 43 |
| e | -10 | 18 | 45 | 68 |
| f | 20 | 46 | 70 | 89 |

C) ELEMENTARY ANALYSIS of panel B -- by means

| | | | | | eff | fit |
|---|---|---|---|---|---|---|
| a | 12 | 0 | -6 | -7 | -64 | -75 |
| b | 3 | -1 | -2 | 0 | -42 | -53 |
| c | -3 | -1 | 1 | 4 | -16 | -27 |
| d | -5 | -1 | 2 | 4 | 13 | 2 |
| e | -4 | 1 | 3 | 1 | 41 | 30 |
| f | 0 | 3 | 2 | -4 | 67 | 56 |
| eff | -36 | -13 | 12 | 27 | -11 | 00 |
| fit | -47 | -24 | 1 | 26 | 00 | 11 |

D) The DATA of panel A in f-LOGS -- unit 0.01

| | | | | |
|---|---|---|---|---|
| a | -255 | -137 | -84 | -48 |
| b | -131 | -81 | -46 | -16 |
| c | -78 | -44 | -14 | 14 |
| d | -41 | -12 | 16 | 46 |
| e | -10 | 18 | 48 | 84 |
| f | 20 | 51 | 85 | 142 |

exhibit 9 (Cont'd.) - 2

E) ELEMENTARY ANALYSIS OF panel D -- by means

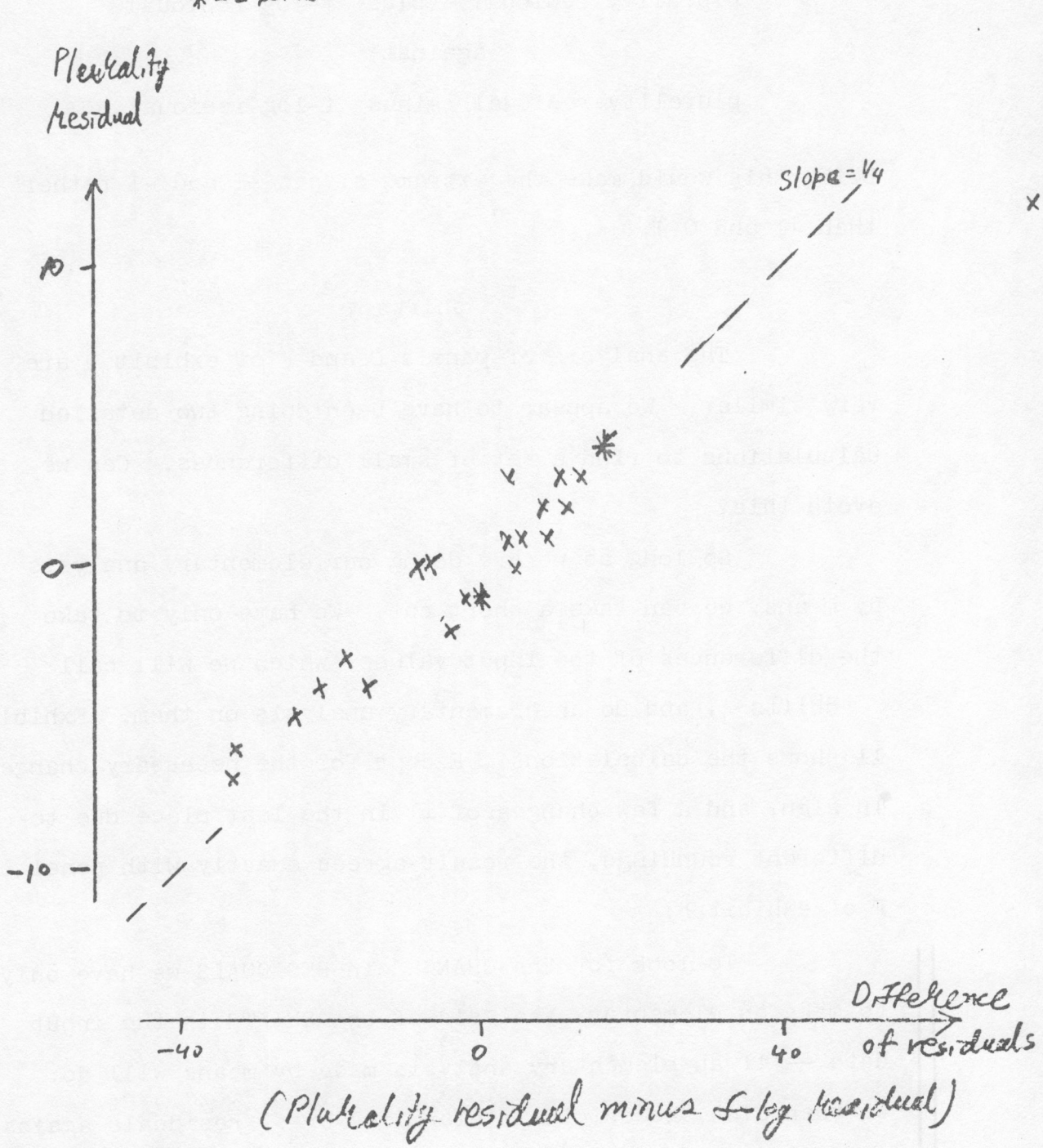
| | | | | | eff | fit |
|---|---|---|---|---|---|---|
| a | -61 | 8 | 27 | 26 | -111 | -131 |
| b | 0 | 1 | 2 | -5 | -48 | -68 |
| c | 15 | 0 | -4 | -13 | -10 | -30 |
| d | 20 | 0 | -6 | -13 | 22 | 2 |
| e | 18 | -3 | -7 | -8 | 55 | 35 |
| f | 9 | -9 | -9 | 11 | 94 | 55 |
| eff | -63 | -14 | 20 | 57 | -20 | 00 |
| fit | -83 | -34 | 0 | 37 | 00 | 20 |

F) Panel C RESIDUALS MINUS panel E RESIDUALS

| | | | | |
|---|---|---|---|---|
| a | 73 | -8 | -33 | -33 |
| b | 3 | -2 | -4 | 5 |
| c | -18 | -1 | 5 | 17 |
| d | -25 | -1 | 8 | 17 |
| e | -22 | 4 | 10 | 9 |
| f | -9 | 12 | 11 | -15 |

exhibit 10 of chapter 28 -- casual example

Panel C residuals plotted against panel F differences
-- both from exhibit 9
(Data additive in f-roots)



(If it were not for the particular way that our calculations will develop, we would have chosen to plot

plurality residual plus f-log residual

against

plurality residual minus f-log residual.

Doing this would make the extreme slopes +1 and -1 rather than +1 and 0.)

* shifts *

The analysis of panels C and F of exhibit 9 are very similar. We appear to have been doing two detailed calculations to find a set of small differences. Can we avoid this?

So long as we are doing our elementary analysis by means, we can take a short cut. We have only to take the differences of the input values, which we will call shifts , and do an elementary analysis on them. Exhibit 11 shows the calculations. Except for the necessary change in sign, and a few changes of ±1 in the last place due to different roundings, the result agrees exactly with panel F of exhibit 9 .

To look for the CHANGES in RESIDUALS we have only to make an elementary analysis of the SHIFTS in the input data -- if an elementary analysis made by means will do. We then plot the original (plurality or %) residuals against these shifts.

exhibit 11 of chapter 28 -- casual example

Analysis of Shifts from Pluralities to f-logs

A) Panel D of exhibit 9 MINUS panel B of same

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| a | -156 | -49 | -15 | -3 |
| b | -47 | -14 | -3 | 0 |
| c | -12 | -3 | 0 | 0 |
| d | -2 | 0 | 0 | 3 |
| e | 0 | 0 | 3 | 16 |
| f | 0 | 5 | 15 | 53 |

B) ELEMENTARY ANALYSIS of panel A -- by means, unit = 0.01

| | | | | | ff | |
|---|---|---|---|---|---|---|
| a | -73 | -8 | 32 | 33 | -47 | -56 |
| b | -4 | 3 | 4 | -4 | -7 | -16 |
| c | 19 | 2 | -5 | -16 | 5 | -4 |
| d | 25 | 1 | -9 | -17 | 9 | 0 |
| e | 22 | -4 | -11 | -9 | 14 | 5 |
| f | 9 | -12 | -12 | 15 | 27 | 18 |
| ff | -27 | -1 | 9 | 20 | -9 | 0 |
| | -36 | -10 | 0 | 11 | 0 | 9 |

Note: Agreement to ±1 with the negative of panel F of exhibit 9.

This short cut need not work for resistant elementary analysis. The basic idea is repeating the analysis and looking at the differences -- only making our elementary analysis using means guarantees a simple short cut.

* finding the shift *

To find the shifts for analysis we can read differences across columns from exhibit 1. This is usually quite good enough.

Occasionally we have data on which we would like to do arithmetic in more detail. Such data is likely to be in %. Recalling exhibit 3, we see that the most useful shifts are then

(f-root in %) - (%)

and

(f-log in %) - (%)

Exhibit 12 provides a table of the former in 0.1% steps. A break table is given for % between 15 and 85. A tail table is given for the remaining cases. Exhibits 13 and 14 give, respectively, the break and tail tables for

(f-log in %) - (%).

We are now ready to find shifts when and as desired.

exhibit 12 of chapter 28 -- shifts tabled

Values of (f-root in %) - (%)

All units = 0.1%

Break Table (15% to 85%)

| - Break | | + Break |
|---|---|---|
| 401 | 0 | 599 |
| 357 | 1 | 643 |
| 332 | 2 | 668 |
| 312 | 3 | 688 |
| 298 | 4 | 702 |
| 284 | 5 | 716 |
| 273 | 6 | 727 |
| 264 | 7 | 736 |
| 253 | 8 | 747 |
| 244 | 9 | 756 |
| 237 | 10 | 763 |
| 229 | 11 | 771 |
| 223 | 12 | 777 |
| 215 | 13 | 785 |
| 210 | 14 | 790 |
| 203 | 15 | 797 |
| 199 | 16 | 801 |
| 193 | 17 | 807 |
| 189 | 18 | 811 |
| 183 | 19 | 817 |
| 180 | 20 | 820 |
| 174 | 21 | 826 |
| 171 | 22 | 829 |
| 166 | 23 | 834 |
| 163 | 24 | 841 |
| 159 | 25 | 844 |
| 156 | 26 | 849 |
| 1 | 27 | 851 |
| +9 | 28 | |

Note: Matches go to even number

Example: 31.1% 311 between 298 and 312, hence 3, which is to be read as a shift of -0.3%.

(0% to 15%) Conventional table (85% to 100%)

% to (f-root in %)

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| 85 | 28 | 28 | 29 | 29 | 29 | 30 | 30 | 30 | 31 | 31 | 31 | 35 | 38 | 43 | 47 | 52 | 58 | 65 | 72 | 81 | 91 | 104 | 120 | 143 | 207 |
| 86 | 31 | 32 | 32 | 32 | 33 | 33 | 33 | 34 | 34 | 34 | | | | | | | | | | | | | | | |
| 87 | 35 | 35 | 35 | 36 | 36 | 36 | 37 | 37 | 38 | 38 | | | | | | | | | | | | | | | |
| 88 | 38 | 39 | 39 | 40 | 40 | 40 | 41 | 41 | 42 | 42 | | | | | | | | | | | | | | | |
| 89 | 43 | 43 | 43 | 44 | 44 | 45 | 45 | 46 | 46 | 47 | | | | | | | | | | | | | | | |
| 90 | 47 | 48 | 48 | 49 | 49 | 50 | 50 | 51 | 51 | 52 | | | | | | | | | | | | | | | |
| 91 | 52 | 53 | 54 | 54 | 55 | 55 | 56 | 56 | 57 | 58 | | | | | | | | | | | | | | | |
| 92 | 58 | 59 | 59 | 60 | 61 | 61 | 62 | 63 | 63 | 64 | | | | | | | | | | | | | | | |
| 93 | 65 | 66 | 66 | 67 | 68 | 68 | 69 | 70 | 71 | 72 | | | | | | | | | | | | | | | |
| 94 | 72 | 73 | 74 | 75 | 76 | 77 | 77 | 78 | 79 | 80 | | | | | | | | | | | | | | | |
| 95 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | | | | | | | | | | | | | | | |
| 96 | 91 | 93 | 94 | 95 | 96 | 97 | 99 | 100 | 101 | 102 | | | | | | | | | | | | | | | |
| 97 | 104 | 105 | 107 | 108 | 110 | 111 | 113 | 115 | 116 | 118 | | | | | | | | | | | | | | | |
| 98 | 120 | 122 | 124 | 126 | 128 | 130 | 132 | 135 | 137 | 140 | | | | | | | | | | | | | | | |
| 99 | 143 | 146 | 149 | 152 | 156 | 160 | 165 | 170 | 177 | 185 | | | | | | | | | | | | | | | |
| | 00 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | | | | | | | | | | | | | | | |

For % >85, use left (in %) plus top (in 0.1%) scales, + sign.

For % <15, use right (in %) plus bottom (in 0.1%) scales, - sign.

Examples: 98.7% gives 135, a shift of +13.5%.
11.1% gives 42, a shift of -4.2%.

exhibit 13 of chapter 28 -- shifts tabled

(f-log in %) - %

Break table for 15% to 85%

All units are 0.1%

| 24.7% to 75.3% | | | 19.4% to 24.7%<br>75.3% to 80.6% | | | 15.8% to 19.4%<br>80.6% to 84.2% | | | 15.0% to 15.8%<br>84.2% to 85.0% | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| - Break | | + Break | - Break | | + Break | - Break | | + Break | - Break | | + Break |
| 429 | 0 | 571 | 247 | 25 | 753 | 194 | 50 | 806 | 159 | 75 | 841 |
| 396 | 1 | 604 | 245 | 26 | 755 | 191 | 51 | 809 | 157h | 76 | 842h |
| 378 | 2 | 622 | 241 | 27 | 759 | 191 | 52 | 809 | 156h | 77 | 843h |
| 364 | 3 | 636 | 240 | 28 | 760 | 188 | 53 | 812 | 155h | 78 | 844h |
| 353 | 4 | 647 | 236 | 29 | 764 | 187 | 54 | 813 | 154h | 79 | 845h |
| 342 | 5 | 658 | 234 | 30 | 766 | 185 | 55 | 815 | 153h | 80 | 846h |
| 335 | 6 | 665 | 231 | 31 | 769 | 184 | 56 | 816 | 152 | 81 | 848 |
| 326 | 7 | 674 | 230 | 32 | 770 | 182 | 57 | 818 | 151h | 82 | 848h |
| 320 | 8 | 680 | 226 | 33 | 774 | 181 | 58 | 819 | 150h | 83 | 849h |
| 313 | 9 | 687 | 225 | 34 | 775 | 179 | 59 | 821 | 149h | 84 | 850h |
| 308 | 10 | 692 | 222 | 35 | 778 | 178 | 60 | 822 | | | |
| 301 | 11 | 699 | 220 | 36 | 780 | 176 | 61 | 824 | | | |
| 297 | 12 | 703 | 217 | 37 | 783 | 176 | 62 | 824 | | | |
| 291 | 13 | 709 | 216 | 38 | 784 | 173 | 63 | 827 | | | |
| 288 | 14 | 712 | 213 | 39 | 787 | 173 | 64 | 827 | | | |
| 282 | 15 | 717 | 212 | 40 | 788 | 171 | 65 | 829 | | | |
| 278 | 16 | 721 | 209 | 41 | 791 | 170 | 66 | 830 | | | |
| 274 | 17 | 726 | 208 | 42 | 792 | 168 | 67 | 832 | | | |
| 271 | 18 | 729 | 205 | 43 | 795 | 168 | 68 | 832 | | | |
| 266 | 19 | 734 | 204 | 44 | 796 | 166 | 69 | 834 | | | |
| 264 | 20 | 736 | 201 | 45 | 799 | 165 | 70 | 835 | | | |
| 259 | 21 | 741 | 201 | 46 | 799 | 163 | 71 | 837 | | | |
| 257 | 22 | 743 | 198 | 47 | 802 | 163 | 72 | 837 | | | |
| 253 | 23 | 747 | 197 | 48 | 803 | 165 | 73 | 839 | | | |
| 251 | 24 | 749 | 194 | 49 | 806 | 166 | 74 | 840 | | | |
| 247 | 25 | 753 | 194 | 50 | 806 | 168 | 75 | 841 | | | |

Examples:

27.8% is 278 which gives 16, rounding to evens, which is -1.6%.
63.7% is 637 which gives 4 or +0.4%.
77.3% is 773 which gives 33 or +3.3%.

exhibit 14 of chapter 28 -- shifts tabled

(f-log in %) - (%)

Table for 0% to 15% and 85% to 100%

All units are 0.1%

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 85 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 14 |
| 86 | 94 | 95 | 96 | 97 | 98 | 99 | 100 | 102 | 103 | 104 | 105 | 13 |
| 87 | 105 | 106 | 108 | 109 | 110 | 111 | 113 | 114 | 115 | 117 | 118 | 12 |
| 88 | 118 | 119 | 121 | 122 | 124 | 125 | 127 | 128 | 130 | 131 | 133 | 11 |
| 89 | 133 | 134 | 136 | 137 | 139 | 141 | 142 | 144 | 146 | 148 | 149 | 10 |
| 90 | 149 | 151 | 152 | 155 | 157 | 158 | 160 | 162 | 164 | 166 | 168 | 9 |
| 91 | 168 | 170 | 173 | 175 | 177 | 179 | 181 | 184 | 186 | 188 | 191 | 8 |
| 92 | 191 | 193 | 195 | 198 | 200 | 203 | 206 | 208 | 211 | 214 | 217 | 7 |
| 93 | 217 | 219 | 222 | 225 | 228 | 231 | 235 | 238 | 241 | 244 | 248 | 6 |
| 94 | 248 | 251 | 255 | 258 | 262 | 266 | 270 | 274 | 278 | 282 | 286 | 5 |
| 95 | 286 | 290 | 295 | 299 | 304 | 309 | 314 | 319 | 324 | 329 | 334 | 4 |
| 96 | 334 | 340 | 346 | 352 | 358 | 364 | 371 | 377 | 384 | 392 | 399 | 3 |
| 97 | 399 | 407 | 415 | 423 | 432 | 441 | 450 | 460 | 471 | 481 | 493 | 2 |
| 98 | 493 | 505 | 518 | 531 | 546 | 561 | 578 | 595 | 615 | 636 | 659 | 1 |
| 99 | 659 | 684 | 713 | 746 | 783 | 828 | 883 | 954 | * | * | - | 0 |
| | (10) | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | |

For 0% to 15%, use right (in %) plus bottom (in 0.1%) scales, - sign.

For 85% to 100%, use left (in %) plus top (in 0.1%) scales, + sign.

* 998 gives 1055, 999 gives 1227.

Examples: 92.45% gives 202, which is 20.2%.
7.83% gives 194, which is -19.4%.

28D. Mountains and mice (back to Nebraska)

Now we are ready to look at two-way tables of counted data, let us turn back to chapter 17 and the vote of 10 Nevada counties for president in 12 elections. We looked at the conventional diagnostic plots in chapter 17, they told us that we did not need to replace %D by $\sqrt{\%D}$ or log %D . We now know that they did essentially nothing to tell us whether we should replace %D by f-root D in % or by f-log D in % .

Exhibit 15 gives the calculations for shifts to f-logs for the table of third residuals in exhibit 15 of chapter 17. We find in turn:

- the shifts.
- their elementary analysis.
- a display of their residuals.
- end "thirds" of these shift residuals.
- which third residuals match these end "thirds".
- trimeans of these matching input data and of the shift residuals they match.

As pointed out in panel E, the indications are that additivity would be oh so slightly improved if we went to a form of expression on the opposite side of % from f-logs -- to a form with slightly shrunken ends rather than to one with stretched ends.

As a partial check, exhibit 16 repeats the calculation, using the shifts to f-roots in % .

exhibit 15 of chapter 28 -- Nebraska residuals

Diagnosis of the third residuals of exhibit 15
of chapter 17 using the tables of exhibits 13 and 14

A) The SHIFTS -- exhibit 3 of chapter 17 and exhibit 13 above
unit = 0.1%

| | '20 | | '28 | | '36 | | '44 | | '52 | | '60 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I2 | -3 | -3 | -7 | 4 | 3 | 0 | 1 | 0 | -19 | -12 | -7 | 0 |
| H3 | -16 | -15 | -18 | 2 | 0 | -6 | -14 | -2 | -33 | -15 | -20 | 0 |
| G4 | -12 | -6 | -24 | 2 | 0 | -6 | -10 | -2 | -45 | -29 | -19 | 0 |
| G6 | -10 | -7 | -18 | 2 | 0 | -2 | -2 | 0 | -33 | -12 | -6 | 0 |
| H7 | -15 | -7 | -14 | 0 | 1 | -2 | -2 | 0 | -13 | -7 | -4 | 0 |
| G7 | -19 | -9 | -7 | 3 | 0 | -2 | -2 | 0 | -11 | -6 | -4 | 0 |
| E7 | 0 | 0 | 0 | 61 | 16 | 0 | 0 | 3 | -1 | 0 | 0 | 6 |
| G8 | -9 | -5 | -11 | 1 | 0 | -6 | -6 | -2 | -28 | -25 | -18 | 0 |
| E8 | -10 | -7 | -12 | 3 | 0 | -6 | -6 | -2 | -33 | -15 | -10 | 0 |
| E9 | -6 | -6 | -8 | 4 | 0 | -2 | -2 | 1 | -31 | -9 | -7 | 0 |

Examples: -3 ('20 and I2) comes from 37.3 via exhibit 13,

61 ('32 and E7) comes from 82.3 via exhibit 13,

For the observed %'s of 37.3, 82.3, etc. see exhibit 3
of chapter 17 at page 17-5.

B) ELEMENTARY ANALYSIS of panel A

| | | | | | | | | | | | | | eff | fit |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I2 | 5 | 1 | 3 | -6 | -1 | 1 | 4 | -2 | 4 | -1 | 1 | -3 | 2 | -4 |
| H3 | -1 | -4 | -1 | -1 | 3 | 2 | -4 | 3 | -3 | 3 | -5 | 4 | -5 | -11 |
| G4 | 4 | 6 | -6 | 0 | 4 | 3 | 1 | 4 | -14 | -10 | -3 | 5 | -6 | -12 |
| G6 | 1 | 0 | -5 | -5 | -1 | 2 | 4 | 1 | -7 | 2 | 5 | -2 | -1 | -7 |
| H7 | -6 | -2 | -3 | -9 | -2 | 0 | 2 | -1 | 11 | 5 | 5 | -2 | 1 | -5 |
| G7 | -10 | -4 | 4 | -6 | -3 | 0 | 2 | -1 | 13 | 6 | 5 | -2 | 1 | -5 |
| E7 | -3 | -7 | -1 | 40 | 1 | -10 | -8 | -10 | 11 | 0 | -3 | -8 | 13 | 7 |
| G8 | 4 | 4 | 4 | -4 | 1 | 0 | -3 | 1 | 0 | -9 | -5 | 2 | -3 | -9 |
| E8 | 2 | 1 | 2 | -3 | 0 | -1 | 1 | 0 | -6 | 0 | 2 | 1 | -2 | -8 |
| E9 | 3 | -1 | 3 | -5 | -3 | 0 | 2 | -1 | -7 | 3 | 2 | 0 | 1 | -5 |
| eff | -4 | -2 | -6 | 14 | 8 | 3 | 1 | 6 | -19 | -7 | -4 | 5 | -6 | 0 |
| fit | -10 | -8 | -12 | 8 | 2 | -3 | -5 | 0 | -25 | -13 | -10 | 1 | 0 | 6 |

exhibit 15 (Cont'd.) - 2

C) Panel B RESIDUALS displayed and END "QUARTERS" found

| | | | |
|---|---|---|---|
| | to | -3 | |
| | | -2 | |
| | 5 | -1 | 00040 |
| | 17 | -0 | 676696876798 |
| | 22 | -5 | |
| | 26 | -4 | |
| * | 36 | -3 | ⊠ |
| | 42 | -2 | |
| | 54 | -1 | ⊠ |
| | (12) | 0 | ⊠ |
| | 54 | 1 | ⊠ |
| | 42 | 2 | ⊠ |
| | 31 | 3 | □ |
| * | 23 | 4 | ⊠ |
| | 12 | 5 | |
| = | 6 | 0 | 66 |
| | 4 | 1 | 131 |
| | | 2 | |
| | | 3 | |
| | 1 | 4 | 0 |

end quarters

count 26, hinges at depth 7
Hinges: -8 and -5
Median: -6
Trimean: -6

count 23, hinges at depth 6
Hinges: +4 and +6
Median: +5
Trimean: +5

*These cuts are as close to quarters of 54 + 12 + 54 = 120 as we can come.

exhibit 15 (Cont'd.) - 3

D) VALUES of panel B RESIDUALS, TAGS, VALUES of THIRD RESIDUALS (from exhibit 15 of chapter 17)

low "quarter"

| | | |
|---|---|---|
| -14 | G4/52 | 0 |
| -10 | G7/20 | -20 |
| -10 | E7/40 | 4 |
| -10 | E7/48 | 26 |
| -10 | G4/56 | -4 |
| -9 | H7/32 | -32 |
| -9 | G8/56 | -1 |
| -8 | E7/44 | 22 |
| -8 | E7/64 | 18 |
| -7 | E7/24 | 14 |
| -7 | G6/52 | -28 |
| -7 | E9/52 | -24 |
| -6 | H7/20 | 0 |
| -6 | G4/28 | -36 |
| -6 | I2/32 | -21 |
| -6 | G7/32 | 3 |
| -6 | E8/52 | 0 |
| -5 | G6/28 | -28 |
| -5 | G6/32 | -11 |
| -5 | E9/32 | 9 |
| -5 | H3/60 | -17 |
| -5 | G8/60 | 4 |
| -4 | H3/24 | 11 |
| -4 | G7/24 | -9 |
| -4 | G8/32 | 14 |
| -4 | H3/44 | -22 |

| | |
|---|---|
| -3 | 26 |
| -2 | 084182 |
| -1 | 17 |
| -0 | 04109 |
| 0 | 40394 |
| 1 | 8414 |
| 2 | 62 |
| 3 | |

count 26,
Hinges: -21 and +9
Median: -0h
Trimean: -3

high "quarter"

| | | |
|---|---|---|
| 40 | E7/32 | -2 |
| 13 | G7/52 | 19 |
| 11 | H7/52 | 8 |
| 11 | E7/52 | -11 |
| 6 | G4/24 | 11 |
| 6 | G7/56 | -3 |
| 5 | I2/20 | 20 |
| 5 | H7/56 | -7 |
| 5 | G6/60 | 12 |
| 5 | H7/60 | -12 |
| 5 | G7/60 | -17 |
| 5 | G4/64 | 8 |
| 4 | G4/20 | -1 |
| 4 | G8/20 | 10 |
| 4 | G8/24 | 14 |
| 4 | G7/28 | 32 |
| 4 | G8/28 | 13 |
| 4 | G4/36 | 6 |
| 4 | I2/44 | 14 |
| 4 | G6/44 | 22 |
| 4 | G4/48 | 1 |
| 4 | I2/52 | 11 |
| 4 | H3/64 | -17 |

| | |
|---|---|
| -2 | |
| -1 | 1277 |
| -0 | 2371 |
| 0 | 8861 |
| 1 | 91204341 |
| 2 | 02 |
| 3 | 2 |

count 23,
Hinges: -3 and +13
Median: +1
Trimean: +3

E) DIAGNOSTIC TILT

$$-\frac{+3-(-3)}{+5-(-6)} = -\frac{6}{11} = -0.6$$

Since residuals shifting + already had a + trimean, and those shifting - a - trimean, going from % toward either f-roots or f-logs would make the third residuals bigger. Thus, if anything, this analysis suggests shrinking together rather than stretching out for the expression as we move reasonably well away from 50%.

exhibit 16 of chapter 28 -- Nebraska residuals

Diagnosis of the third residuals of chapter 17

using the shifts of exhibit 12

A) The SHIFTS -- unit 0.1%

| | '20 | | '28 | | '36 | | '44 | | '52 | | '60 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I2 | -2 | -1 | -2 | 1 | 1 | 0 | 0 | 0 | -7 | -4 | -3 | 0 |
| H3 | -6 | -5 | -6 | 1 | 0 | -3 | -5 | -1 | -12 | -6 | -7 | 0 |
| G4 | -4 | -2 | -9 | 1 | 0 | -3 | -4 | -1 | -16 | -10 | -7 | 0 |
| G6 | -3 | -3 | -6 | 1 | 0 | -1 | -1 | 0 | -13 | -5 | -2 | 0 |
| H7 | -5 | -3 | -5 | 0 | 0 | -1 | -1 | 0 | -5 | -3 | -2 | 0 |
| G7 | -7 | -3 | -3 | 1 | 0 | 0 | -1 | 0 | -4 | -2 | -1 | 0 |
| E7 | 0 | 0 | 0 | 21 | 6 | -2 | 0 | 1 | 0 | 0 | 0 | 2 |
| G8 | -3 | -2 | -4 | 0 | 0 | -2 | -4 | -1 | -10 | -9 | -6 | 0 |
| E8 | -4 | -3 | -4 | 1 | 0 | -2 | -2 | -1 | -11 | -6 | -4 | 0 |
| E9 | -2 | -2 | -3 | 1 | 0 | -1 | -1 | 0 | -11 | -4 | -3 | 0 |

B) ELEMENTARY ANALYSIS of Panel A -- by means

| | | | | | | | | | | | | | eff | fi |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I2 | 1 | 0 | 1 | -1 | -1 | 0 | 1 | -1 | 1 | 0 | 0 | -1 | 1 | -1 |
| H3 | 0 | -1 | 0 | 0 | 1 | 0 | -1 | 1 | -1 | 1 | -1 | 2 | -2 | -5 |
| G4 | 2 | 2 | -3 | 0 | 1 | 0 | 0 | 1 | -5 | -3 | -1 | 2 | -2 | -4 |
| G6 | 1 | -1 | -2 | -2 | -1 | 0 | 1 | 0 | -4 | 0 | 2 | 0 | 0 | -2 |
| H7 | -1 | -1 | -1 | -3 | -1 | 0 | 1 | 0 | 4 | -2 | 2 | 0 | 0 | -2 |
| G7 | -3 | -1 | 1 | -2 | -1 | 1 | 1 | 0 | 5 | -3 | 3 | 0 | 0 | -2 |
| E7 | -1 | -3 | -1 | 13 | 0 | -6 | -3 | -4 | 4 | 0 | -1 | -4 | 5 | 3 |
| G8 | 2 | 1 | 1 | -2 | 0 | 0 | -1 | 0 | 0 | -3 | -1 | 0 | -1 | -3 |
| E8 | 1 | 0 | 1 | -1 | 0 | 0 | 1 | 0 | -1 | 0 | 0 | 0 | -1 | -3 |
| E9 | 2 | 0 | 1 | -2 | -1 | 6 | 1 | 0 | -2 | 1 | 1 | 0 | 0 | -2 |
| eff | -2 | 0 | -2 | 5 | 3 | 1 | 0 | 2 | -7 | -3 | -2 | 2 | -2 | 0 |
| fit | -4 | -2 | -4 | 3 | 1 | -1 | -2 | 0 | -9 | -5 | -4 | 0 | 0 | 2 |

C) SHIFT RESIDUALS, and END QUARTERS

| to | | |
|---|---|---|
| 2 | -1 | |
| | -0 | 65 |
| 5 | -4 | . |
| 13 | -3 | □ |
| 20 | -2 | □ |
| 45 | -1 | ⊡ ⊡ .. |
| (38) | 0 | □ ⊡ ⊠ □ |
| 37 | 1 | ⊠ ⊠ .. |
| 13 | 2 | □ |
| 5 | 3 | . |
| 4 | 4 | .. |
| 2 | 0 | 5 |
| 1 | 1 | 3 |

| | | |
|---|---|---|
| (Count) | (20) | (13) |
| Hinges: | -3h and -2 | -4 and -3 |
| Median: | -3 | -3 |
| Trimean: | -3 | -3 |

| | | |
|---|---|---|
| (Count) | (13) | (37) |
| Hinges: | -2 and +4 | +1 and +2 |
| Median: | +2 | +1 |
| Trimean: | +2h | +1 |

exhibit 16 (Cont'd.) - 2

D) VALUES of panel B RESIDUALS, TAGS (from B), VALUES of THIRD RESIDUALS (from exhibit 15 of chapter 17)

low "quarter"

| | | | | |
|---|---|---|---|---|
| -6 | E7/40 | 4 | -3 | 62 |
| -5 | G4/52 | 0 | -2 | 8084 |
| -4 | E7/48 | 26 | -1 | 1 |
| -4 | G6/52 | -28 | -0 | 043197 |
| -4 | E7/64 | 18 | 0 | 439 |
| -3 | G7/20 | -20 | 1 | 84 |
| -3 | E7/24 | 14 | 2 | 62 |
| -3 | G4/28 | -36 | | |
| -3 | H7/32 | -32 | | |
| -3 | E7/44 | 22 | | |
| -3 | G4/56 | -4 | | |
| -3 | G7/56 | -3 | | |
| -3 | G8/56 | -1 | | |
| -2 | G6/28 | -28 | | |
| -2 | G6/32 | -11 | | |
| -2 | G7/32 | 3 | | |
| -2 | G8/32 | -9 | | |
| -2 | E9/32 | 9 | | |
| -2 | E9/52 | -24 | | |
| -2 | H7/56 | -7 | | |

(count 20,50,5h,10h)
Hinges: -22 and +6h
Median: -3h
Trimean: -6

(less -2's, see F)

(count 13,50,4,7)
Hinges: -20 and +14
Median: -1
Trimean: -2

high "quarter"

| | | | | |
|---|---|---|---|---|
| 13 | E7/32 | -2 | -2 | |
| 5 | G7/52 | 19 | -1 | 1727 |
| 4 | H7/52 | 8 | -0 | 21 |
| 4 | E7/52 | -11 | 0 | 88 |
| 3 | G7/60 | -17 | 1 | 9012 |
| 2 | G4/20 | -1 | 2 | 1 |
| 2 | G8/20 | 10 | | |
| 2 | E9/20 | 21 | | |
| 2 | G4/24 | 11 | | |
| 2 | G6/60 | 12 | | |
| 2 | H7/60 | -12 | | |
| 2 | H3/64 | -17 | | |
| 2 | G4/64 | 8 | | |

(Count 13,80,4,7)
Hinges: -11 and +11
Median: +8
Trimean: +4

(with +1's, see F)

(Count 37,50,10,19)
Hinges: -2 and +14
Median: +8
Trimean: +7

Note: If we go as close to 1/3 at each end as we can without breaking ties, we would go to -2 (20 residuals) and +1 (37 residuals). It is natural to feel that +1 is too close to zero to be reasonable. This leads to using -2 (20 residuals) and +2 (13 residuals). Some feel that better balance should be sought. This would lead to -3 (13 residuals) and +2 (13 residuals).

E) DIAGNOSTIC TILTS

$$(20 \text{ vs } 37) - \frac{+7 - (-6)}{+1 - (-3)} = - \frac{13}{4} = - 3.25$$

$$(20 \text{ vs } 13) = \frac{+4 - (-6)}{+2 - (-3)} = - \frac{10}{5} = - 2.0$$

$$(13 \text{ vs } 13) = \frac{+4 - (-2)}{+2 - (-3)} = - \frac{6}{5} = - 1.2$$

Again suggests going in the direction away from f-roots and f-logs.

exhibit 16 (Cont'd.) - 3

F) ALTERNATE STEM-AND-LEAVES FOR D

| to | | |
|---|---|---|
| 2 | -3 | 62 |
| 4 | -2 | 80 |
| ( ) | -1 | |
| 8 | -0 | 0431 |
| 5 | 0 | 4 |
| 4 | 1 | 84 |
| 2 | 2 | 62 |

(Count 13,50,4,7)
Hinges: -20 and +14
Median: -1
Trimean: -2

| | |
|---|---|
| I2/20 | 20 |
| G6/20 | 19 |
| E8/20 | 11 |
| G8/24 | 14 |
| I2/28 | -12 |
| G7/28 | 32 |
| G8/28 | 13 |
| E8/28 | 7 |
| E9/28 | 9 |
| H3/36 | 26 |
| G4/36 | 6 |
| G7/40 | -4 |
| I2/44 | 8 |
| G6/44 | 22 |
| H7/44 | 8 |
| G7/44 | -2 |
| E8/44 | -3 |
| E9/44 | 20 |
| H3/48 | 3 |
| G4/48 | 1 |
| I2/52 | -6 |
| H3/56 | 24 |
| E9/56 | 12 |
| E9/60 | 7 |

| to | | |
|---|---|---|
| 5 | -1 | 17272 |
| 11 | -0 | 214236 |
| 21 | 0 | 8879688317 |
| 16 | 1 | 901291432 |
| 7 | 2 | 106204 |
| 1 | 3 | 2 |

(Count 37,50,10,19)
Hinges: -2 and +14
Median: +8
Trimean: +7

Where do we stand? A hard look at panel D of either exhibit 15 or exhibit 16 convinces us that we are looking at small differences among widely varying quantities. Against this we have to give some weight to the steadfastness of the sign of the diagnostic tilt. On balance, we ought not to let one example of this sort convince us, but two or three more such easily might.

Taking the sign of the diagnostic tilt seriously, so that we may see what it would mean, we can sum up the possible conclusions from this example as follows:

- if f-logs had been the proper expression, this would have made very little difference -- only one residual of 120 being changed more than 1.4% -- and only nine as much as 1.0%.

- if f-roots had been the proper expression, even less would have changed -- only one residual of 120 being changed more than 0.6% -- and only nine as much as 0.4%.

- as a consequence, evidence on what form of expression would be good is inevitably weak.

- taken at face value, the available evidence suggests neither f-logs or f-roots, but some form of expression deviating from % in the opposite direction.

In view of our success with f-logs in straightening opinion polls on population segments, there must be some

concern about this last result (even if the evidence is weak). Were it to be true, wouldn't this be a little paradoxical? Or would it?

* behavior of a composite *

The key to our concern is the fact that made our Kennedy support example possible -- the electorate is not an unstructured whole. It has segments of quite different character. An election is not like a poll -- the only vote that it is proper for any one to know is a total -- a composite over all the segments. Perhaps we should try to dream up an example of a composite population -- say one where each segment behaves nicely in f-logs.

Exhibit 17 does just this, starting with some precise % corresponding to round f-logs. The composite shows a % vote increasing faster -- say when near 80% -- than the % matched to it at 50%. (This matching could be done by calculating the effects of much smaller changes in conditions than the change from A to B.) The lettered conditions, A,B,C,...,I correspond to equal steps of f-logs in panel A. Hence they correspond, under our assumptions to equal steps in the overall nature of the circumstances and should be reflected by equal steps in the ideal expression.

exhibit 17 of chapter 28 -- composite behavior

Example of composite behavior when all segments behave according to f-logs

A) Some % corresponding to ROUND f-LOGS

| f-log | % | f-log | % |
|---|---|---|---|
| 0 | 50.00 | 0 | 50.00 |
| 0.5 | 73.11 | -0.5 | 26.89 |
| 1.0 | 88.08 | -1.0 | 11.92 |
| 1.5 | 95.26 | -1.5 | 4.74 |
| 2.0 | 98.20 | -2.0 | 1.80 |
| 2.5 | 99.33 | -2.5 | 0.67 |
| 3.0 | 99.77 | -3.0 | 0.23 |
| 3.5 | 99.91 | -3.5 | 0.09 |
| 4.0 | 99.97 | -4.0 | 0.03 |
| 4.5 | 99.99 | -4.5 | 0.01 |
| 5.0 | 100.00 | -5.0 | 0.00 |

B) HYPOTHETICAL 5-SEGMENT COMPOSITE -- entries in f-logs

| Size | Various conditions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| 20,000 | -2 | -1.5 | -1 | -0.5 | 0 | 0.8 | 1 | 1.5 | 2 |
| 10,000 | -1 | - .5 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| 10,000 | 0 | .5 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| 10,000 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 |
| 20,000 | 2 | 2.5 | 3 | 3.5 | 4 | 4.5 | 5 | 5.5 | 6 |

C) Same COMPOSITE -- entries in %

| Size | Various conditions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| 20,000 | 1.80 | 4.74 | 11.92 | 26.89 | 50.00 | 73.11 | 88.08 | 95.26 | 98.20 |
| 10,000 | 11.92 | 26.89 | 50.00 | 73.11 | 88.08 | 95.26 | 98.20 | 99.33 | 99.77 |
| 10,000 | 50.00 | 73.11 | 88.08 | 95.26 | 98.20 | 99.33 | 99.77 | 99.91 | 99.99 |
| 10,000 | 88.08 | 95.26 | 98.20 | 99.33 | 99.77 | 99.91 | 99.99 | 100.00 | 100.00 |
| 20,000 | 98.20 | 99.33 | 99.77 | 99.91 | 99.99 | 100.00 | 100.00 | 100.00 | 100.00 |

exhibit 17 (Cont'd.) - 2

D) Same COMPOSITE -- entries in votes

| Size | Various conditions | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I |
| 20,000 | 360 | 948 | 2,384 | 5,378 | 10,000 | 14,622 | 17,616 | 19,052 | 19,640 |
| 10,000 | 1,192 | 2,689 | 5,000 | 7,311 | 8,808 | 9,526 | 9,820 | 9,933 | 9,977 |
| 10,000 | 5,000 | 7,311 | 8,808 | 9,526 | 9,820 | 9,933 | 9,977 | 9,991 | 9,999 |
| 10,000 | 8,808 | 9,526 | 9,820 | 9,933 | 9,977 | 9,991 | 9,999 | 10,000 | 10,000 |
| 20,000 | 19,640 | 19,866 | 19,954 | 19,982 | 19,998 | 20,000 | 20,000 | 20,000 | 20,000 |
| total | 35,000 | 40,340 | 45,966 | 52,130 | 58,603 | 63,972 | 67,412 | 68,976 | 69,616 |
| as % | 50.00 | 57.63 | 65.67 | 74.47 | 83.72 | 91.39 | 96.30 | 98.54 | 99.45 |
| matched % | 50.00 | 57.58 | 65.16 | 72.74 | 80.32 | 87.90 | 95.48 | 103.06 | 110.64 |
| diff | .00 | +.05 | +.51 | +1.73 | +2.40 | +3.49 | +0.82 | -4.52 | -11.19 |

matched % = equal steps in % matched to "as % "at 50%.

E) NOTE

Differences from A to B to C --- to I are equal steps in f-logs for each segment. Therefore these are equal steps for changing circumstances. If circumstances are additive, the expression that would make the data additive -- the ideal expression -- would assign the %s in the as % line to equally spaced values of this expression.

The easiest way to see that this expression is on the other side of % from f-log -- at least out to 90% or 95% -- is to make a plot. Exhibit 18 does just that. The difficulty of converting the last few people in the hard core eventually brings the ideal expression back up above the % line, but only very close to 100%.

If we have the right mixture of "hard-core Republicans" , "regular Republicans" , "independents", "regular Democrats" , and "hard-core Democrats" , with each of these segments behaving according to f-logs, we can have the composite behave quite unlike f-logs , moving away from % on the side away from f-logs over the range from 15% to 85%.

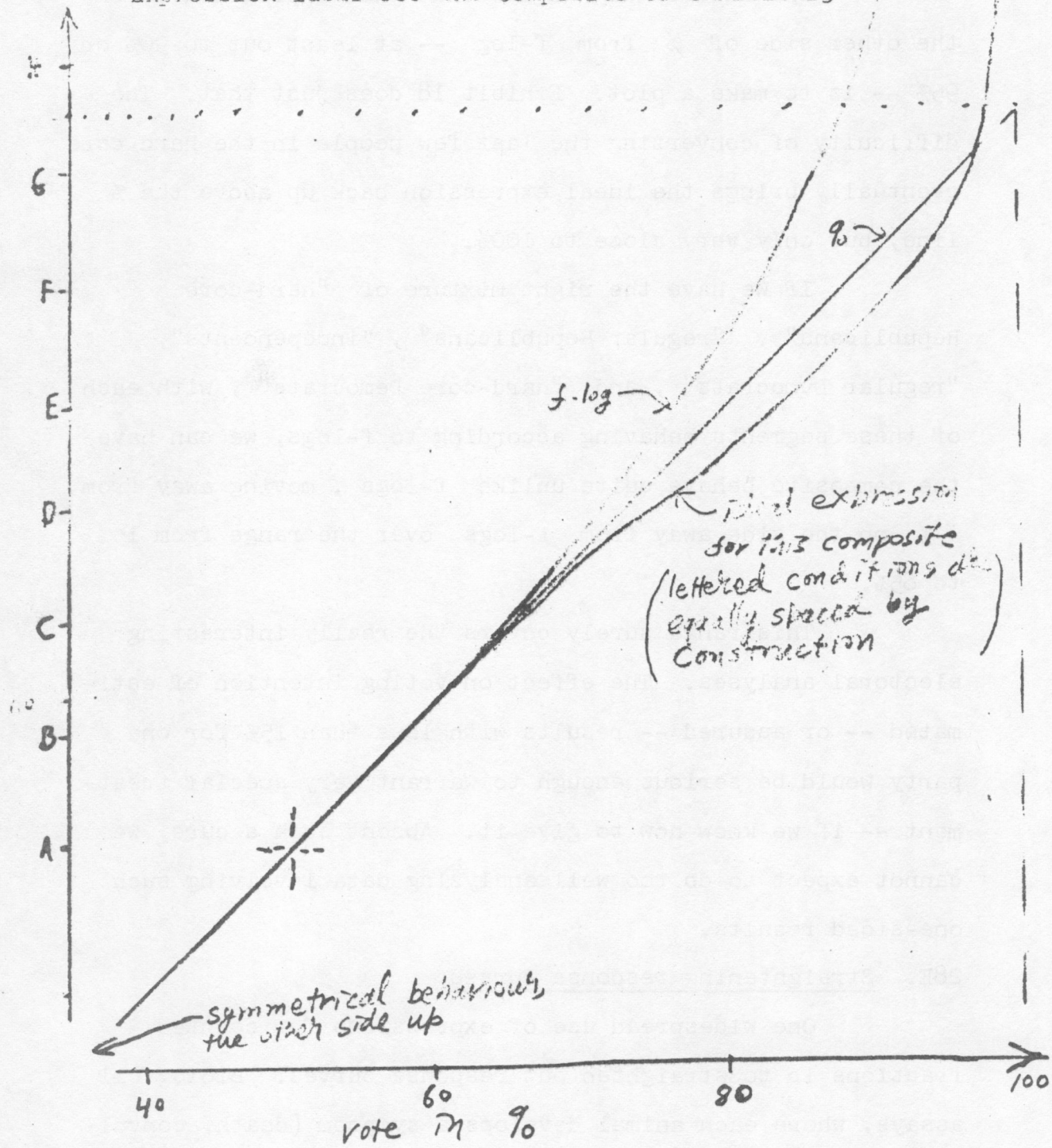
This range surely covers the really interesting electoral analyses. The effect on voting intention of estimated -- or assured -- results with less than 15% for one party would be serious enough to warrant very special treatment -- if we knew how to give it. Absent such a cure, we cannot expect to do too well analyzing data involving such one-sided results.

28E. Straightening response curves

One widespread use of expressions for counted fractions is to straighten out response curves. Biological assays, where each animal develops a symptom (death, convulsion, sleepiness, a disease, etc.) or doesn't, are essential

exhibit 18 of chapter 28 -- composite behavior

Comparison of three expressions: %, f-logs, and the expression ideal for the composite of exhibit 13



to measure the potency of many drugs and vaccines. All that we know of biological responses suggests strongly that if we plot the % responding as a function of dose we should see an s-shaped curve.

Since both looking at residuals and doing arithmetic is simpler when we fit straight lines, s-shaped curves in % suggest strongly using some expression like f-roots or f-logs for the counted fractions

$$\frac{\text{number of animals responding}}{\text{number of animals exposed}}$$

Finding good data that convinces one that such dose-response curves are really s-shaped is not easy. On one hand, any reasonable expressions can be matched to one another near 50%. On the other, economy in both human effort and the number of animals used calls both for avoiding doses with extreme fractions responding and for using as few animals at each dose as will provide an adequate measurement. It almost seems as though data on enough individuals over a wide enough dosage range to be convincing by themselves is confined to tests against insects, where a hundred or more flies per dose is inexpensive and not uncommon.

The same problem arises in connection with the probability of human response. The probability that a dark-adapted human eye will see a faint flash of light is a

definitely s-shaped curve, but not quite a symmetrical one, even if we use the log of the brightness of the flash. (We understand the physics of why this is so.) The probability that a subject will taste an substance present in extreme dilution is also an s-shaped function of the log of the amount present.

One classical set of data was used by Spearman in 1902 in a paper on the analysis of such data. Exhibit 19 gives the numbers; exhibit 20 shows the three strings of points corresponding to %, f-roots and f-logs. The vertical scales, not shown, are uniform. They have been adjusted to make eye comparison of the three curves easy. We see the percent or plurality plot -- circles, to the left -- is definitely s-shaped. The f-log plot -- triangles, to the right -- appears reverse s-shaped. Had we only this data to go on, we would choose f-roots.

The total amount of comparative study that has been made is not as large as one would wish. On balance, it suggests trying f-logs before f-roots, but being glad that choosing one instead of the other rarely makes a crucial difference.

* delicacy and vigor in re-expression *

The last few sections have tried to teach us that re-expression of counted fractions is a delicate thing. Re-expressing amounts was a much more vigorous matter. Why?

exhibit 19 of chapter 28 -- reported loudness

Merkel's data on reported relative loudness

A) DATA

| Actual relative loudness | Number of times reported | | |
|---|---|---|---|
| | Louder | Uncertain | Softer |
| -694 | | 2 | 98 |
| -538 | | 11 | 89 |
| -370 | 11 | 14 | 73 |
| -195 | 28 | 20 | 52 |
| 0 | 48 | 24 | 28 |
| 200 | 71 | 18 | 11 |
| 397 | 83 | 15 | 2 |
| 603 | 91 | 8 | |
| 807 | 95 | 5 | |
| 1021 | 98 | 2 | |
| 1239 | 100 | | |

B) SOURCE

C. Spearman, 1908. The method of 'right and wrong cases' ('constant stimuli') without Gauss's formula. Brit. J. Psychol 2:227 - Table 1 on page 230.

His source: Merkel, 1904. Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik. Page 52ff.

exhibit 20

Three of Merkel's % of times the subject stated that the second sound was louder or equally loud plotted against relative loudness

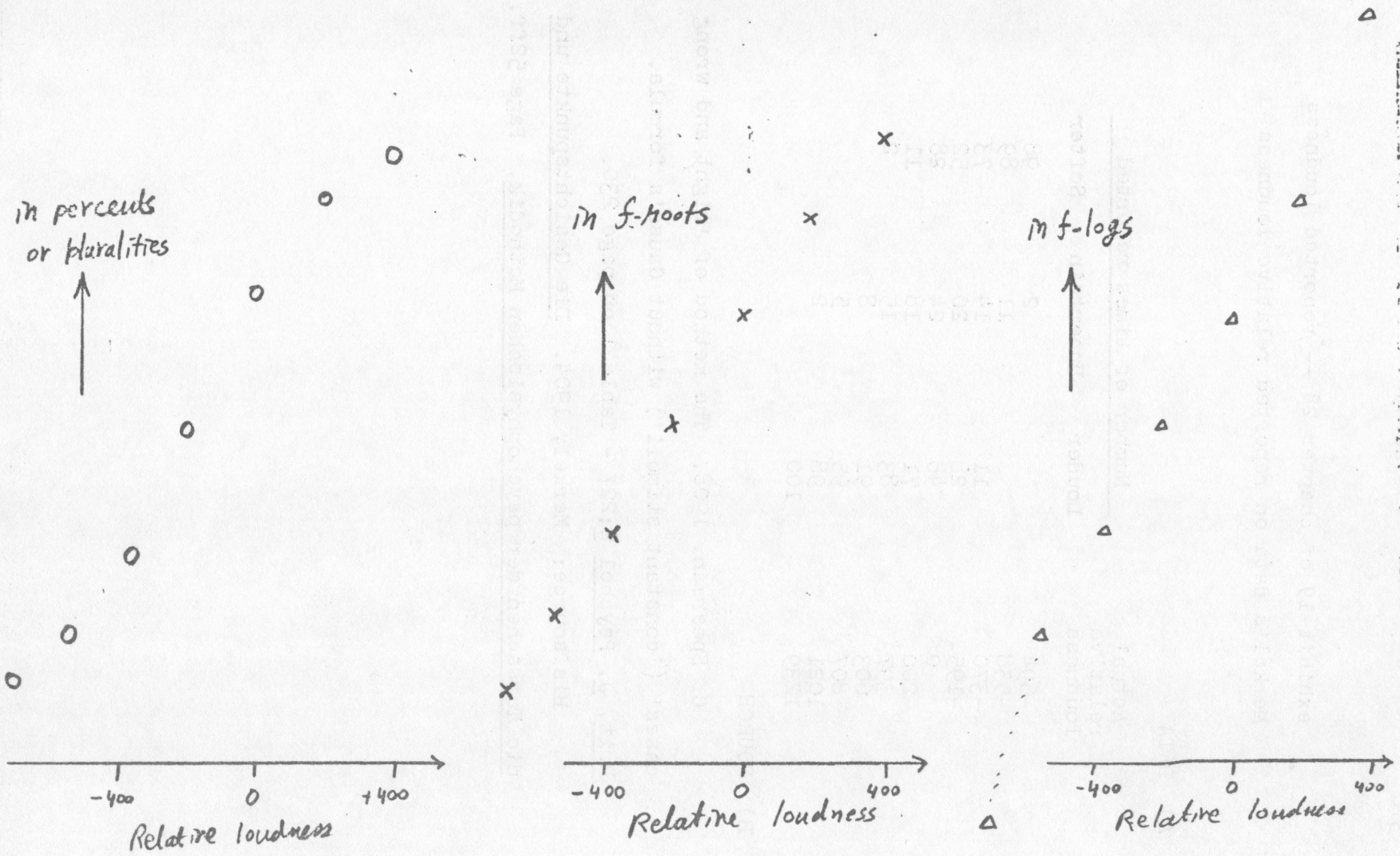





exhibit 20 of chapter 28 -- reported loudness

If we think back to chapter 3, where we began on re-expression, we can recall that if any one of

$$y, \quad \sqrt{y}, \quad \log y, \quad -1/y$$

is a straight-line function of x, and we match another of these expressions to it, the matching expression will pull away from the straight line

on the same side at both ends

This is a vigorous difference, easy to detect.

If we think about exhibit 1 above for a while, we see that the differences among these expressions are more delicate and subtle. If any one expression is straight, any other pulls away in one direction at one end but in the other direction at the other.

Where re-expression of amounts behaves like

$$y + cy^2$$

re-expression of counted fractions behaved like

$$y + by^3$$

The case of dose -- or brightness, concentration or loudness -- and response is a good one to teach us. We begin by adjusting the expression of what is done -- or what are the circumstances. By doing this we can take out simple

curvature up or curvature down. Once this is done, we may have a straight line or we may have an s-shaped curve. If the latter, we make it straighter by re-expressing the counted fraction. This second adjustment is obviously more delicate than the first was.

28F. Double dichotomies

The simplest case of comparing counted fractions comes up when there are two fractions -- and thus a total of 4 counts. This simple case becomes just a little complicated when there is doubt as to which way we ought to take the counted fractions.

To understand how to avoid this complication, it helps to begin with a case where it doesn't occur. Exhibit 21 shows some classical data of Greenwood and Yule on inoculation against -- and attack by -- cholera. Here there is no doubt that we should compare the two fractions attached -- one for these inoculated and the other for those not inoculated.

The symmetrical case, where we are uncertain whether either choice of paired fractions is to be preserved, will be discussed at the beginning of chapter 29x.

exhibit 21 of chapter 28 -- cholera protection

Cholera -- inoculation and attack

A) DATA -- four counts, naturally 2x2

| | Inoculated | Not Inoculated |
|---|---|---|
| Attacked | 3 | 66 |
| Not Attacked | 276 | 473 |
| (sum) | (279) | (539) |

B) % ATTACKED and f-LOG ATTACKED -- f-log in matched form

| | Inoculated | Not Inoculated |
|---|---|---|
| % Attacked | 1.07% | 12.24% |
| f-log | -2.27 | -1.01 |

Difference in f-logs = 1.26

E) SOURCE:

M. Greenwood and G. U. Yule, 1915. The statistics of anti-typhoid and anti-cholera inoculations, and the interpretation of such statistics in general. Proc. Roy. Soc. Medicine 8:113.

Used by: M. G. Kendall. The Advanced Theory of Statistics, 1945, 1:309.

28G. Counts of counted fractions

For most purposes, just how we define f-roots or f-logs matters very little. We can use free starts or leave them alone. Sometimes using free starts can make a difference, however, One case is where we are to use a histogram or rootogram with counted fractions along the base when we have observed some counted fractions that would be 0.00 or 1.00 if we did not use a free start.

A histogram or rootogram for a counted fraction has an interval along the horizontal axis to replace each possible result. The value to be replaced should stand roughly in the middle of the interval. If we do not use a free start, none observed of one or the other kind will provide a counted fraction of either zero or one. Since we have no definitions for f-roots or f-logs beyond zero or one, neither zero or one can then be near the center of an interval.

The easy and effective way to cure this difficulty is to use a free start of 0.5 in each category.

For counts adding to 12, the results are as in exhibit 22. In 189?, Weldon threw 12 dice many times. Some of his results are set down in exhibit 23 and converted to suspended rootograms. The rootograms themselves appear as exhibits 24 and 25. Clearly the fit to the comparison curve --

exhibit 22 of chapter 28 -- fractions of 12

Results of adding free starts of 0.5 to
all counts -- indented lines in parenthesis are cuts

| Actual Counts | Adjusted Counts | Adjusted Fractions | Adjusted f-root | Adjusted f-log |
|---|---|---|---|---|
| | (0 and 13) | (0/13 = 0.000) | (-1.41) | (∞) |
| 0 and 12 | 0.5 and 12.5 | 0.5/13 = 0.038 | -1.11 | -1.61 |
| | (1 and 12) | (1/13 = 0.077) | (-.97) | (-1.24) |
| 1 and 11 | 1.5 and 11.5 | 1.5/13 = 0.115 | -.86 | -1.02 |
| | (2 and 11) | (2/13 = 0.154) | (-.75) | (-.84) |
| 2 and 10 | 2.5 and 10.5 | 2.5/13 = 0.192 | -.65 | -.71 |
| | (3 and 10) | (3/13 = 0.231) | (-.56) | (-.60) |
| 3 and 9 | 3.5 and 9.5 | 3.5/13 = 0.269 | -.47 | -.50 |
| | (4 and 9) | (4/13 = 0.308) | (-.39) | (-.40) |
| 4 and 8 | 4.5 and 8.5 | 4.5/13 = 0.346 | -.31 | -.32 |
| | (5 and 8) | (5/13 = 0.385) | (-.23) | (-.23) |
| 5 and 7 | 5.5 and 7.5 | 5.5/13 = 0.423 | -.16 | -.16 |
| | (6 and 7) | (6/13 = 0.462) | (-.08) | (-.08 |
| 6 and 6 | 6.5 and 7.5 | 6.5/13 = 0.500 | .00 | .00 |
| | (7 and 6) | (7/13 = 0.538) | (.08) | (.08) |
| 7 and 5 | 7.5 and 6.5 | 7.5/13 = 0.577 | .16 | .16 |
| | (8 and 5) | (8/13 = 0.615) | (.23) | (.23) |
| 8 and 4 | 8.5 and 4.5 | 8.5/13 = 0.654 | .31 | .32 |
| | (9 and 4) | (9/13 = .0692) | (.39) | (.40) |
| 9 and 3 | 9.5 and 3.5 | 9.5/13 = 0.711 | .47 | .50 |
| | (10 and 3) | (10/13 = 0.769) | (.56) | (.60) |
| 10 and 2 | 10.5 and 2.5 | 10.5/13 = 0.808 | .65 | .71 |
| | (11 and 2) | (11/13 = 0.846) | (.75) | (.84) |
| 11 and 1 | 11.5 and 1.5 | 11.5/13 = 0.885 | .86 | 1.02 |
| | (12 and 1) | (12/13 = 0.923) | (.97) | (1.24) |
| 12 and 0 | | 12.5/13 = 0.962 | -1.11 | 1.61 |
| | | (13/13 = 1.000) | (-1.41) | (∞) |

exhibit 23 of chapter 28 -- Weldon's dice

Making suspended rootograms for some of

Weldon's dice data

A) 4096 THROWS of 12 DICE -- each 4,5, or 6 called a success

| Successes | No. of Throws | f-roots Cut | f-roots Interval | Throws Interval | $\sqrt{\text{same ratio}}$ |
|---|---|---|---|---|---|
| | | -1.41 | | | |
| 0 | 0 | | .44 | 0 | 0 |
| | | -.97 | | | |
| 1 | 7 | | .22 | 32 | 5h |
| | | -.75 | | | |
| 2 | 60 | | .19 | 316 | 18 |
| | | -.56 | | | |
| 3 | 198 | | .17 | 1162 | 34 |
| | | -.39 | | | |
| 4 | 430 | | .16 | 2684 | 52 |
| | | -.23 | | | |
| 5 | 731 | | .15 | 4878 | 70 |
| | | -.08 | | | |
| 6 | 948 | | .16 | 5925 | 77 |
| | | .08 | | | |
| 7 | 847 | | .15 | 5645 | 75 |
| | | .23 | | | |
| 8 | 536 | | .16 | 3350 | 58 |
| | | .39 | | | |
| 9 | 257 | | .17 | 1512 | 39 |
| | | .56 | | | |
| 10 | 71 | | .19 | 374 | 19h |
| | | .75 | | | |
| 11 | 10 | | .22 | 46 | 6h |
| | | .97 | | | |
| 12 | 0 | | .44 | 0 | 0 |
| | | 1.41 | | | |
| | | | | Sum | 63 |
| | | | | | 391h |
| | | | | | 454h |

All arithmetic to slide-rule accuracy; "h" means "and a half".

exhibit 23 (Cont'd.) - 2

B) FITTING a COMPARISON CURVE -- Hinges are at depth 1024h

| To | Count |
|---|---|
| -.23 | 795 |
| -.08 | 1526 |

hinge near $-.23 + \frac{1024 - 795}{731}(.15) = .18h$

| To | Count |
|---|---|
| +.23 | 872 |
| +.08 | 1719 |

hinge near $+.23 - \frac{1024 - 872}{847}(.15) = .20h$

Note: Because these calculations spread the observations out uniformly over -.23 to -.08 and over +.23 to +.08, while the comparison curve does not spread its observations this way, these hinges will be somewhat too widely separated to give a well fitting comparison curve. A judgement adjustment suffices. We shall make an adjustment of .01h inward in each case, setting the hinges at -.17 and +.19.

C) CALCULATING the COMPARISON CURVE -- for a rootogram

| locs. of centers (and cuts) | | rootogram ordinate | | | |
|---|---|---|---|---|---|
| in f-roots | in hinge units | relative* | fitted | observed** | residual |
| -1.19 | -6.6h | .01 | 1 | 0 | -1 |
| -.86 | -4.8 | .07 | 5h | 5h | 0 |
| -.65h | -3.7 | .21 | 16h | 18 | +1h |
| -.48h | -2.7h | .42 | 33 | 34 | +1 |
| -.31 | -1.7 | .70h | 55h | 52 | -3h |
| -.15h | -0.8h | .92 | 72h | 70 | -2h |
| 0 | -0.0h | 1.00 | 79 | 77 | -2 |
| .15h | 0.8 | .92h | 73 | 75 | +2 |
| .31 | 1.6h | .73 | 57h | 58 | +0h |
| .48h | 2.6h | .45 | 35h | 39 | +3h |
| .65h | 3.6 | 23 | 18 | 19h | +1h |
| .86 | 4.7 | .08 | 6 | 6h | +0h |
| 1.19 | 6.5h | .01 | 1 | 0 | -1 |
| | | (sums) (5.76) to be 459h | (454) ✓ | | |

* from exhibit 13 of chapter 26.

** see panel A.

exhibit 23 (Cont'd.) - 3

D) 4096 THROWS of 12 DICE -- each 6 a success

| Successes | No. of Throws | f-roots* Cut | f-roots* Interval | Throws Interval | $\sqrt{\text{same ratio}}$ |
|---|---|---|---|---|---|
| | | -1.41 | | | |
| 0 | 447 | | .44 | 1016 | 32 |
| | | -.97 | | | |
| 1 | 1145 | | .22 | 5205 | 72 |
| | | -.75 | | | |
| 2 | 1181 | | .19 | 6216 | 79 |
| | | -.56 | | | |
| 3 | 796 | | .17 | 4681 | 68h |
| | | -.39 | | | |
| 4 | 380 | | .16 | 2375 | 48h |
| | | -.23 | | | |
| 5 | 115 | | .15 | 767 | 27h |
| | | -.08 | | | |
| 6 | 24 | | .16 | 150 | 12 |
| | | .08 | | | |
| more** | 8 | | (.15) | (47) | (7) |
| | | (.23) | | | |
| | | | (.16) | (6) | (2h) |
| | | (.39) | | Sum | |
| | | | | Sum | 349 |

All arithmetic to slide-rule accuracy.

E) FITTING of COMPARISON CURVE -- hinges at depth 1024h

| To | Count |
|---|---|
| -.97 | 447 |
| -.75 | 1592 |

hinge near $-.97 + \frac{1024h - 447}{1145}(.23) = -.86$

| To | Count |
|---|---|
| -.39 | 527 |
| -.56 | 1323 |

hinge near $-.39 - \frac{1024h - 527}{796}(.17) = -.50h$

Note: Again we need to adjust. Again a judgement adjustment will be close enough. We take the hinges at -.85 and -.51.

* f-root corresponding to successes/12

**Treated (in parentheses) as 7,1,0,0,0,0.

exhibit 23 (Cont'd.) - 4

F) CALCULATION of COMPARISON CURVE

| location of center | | rootogram ordinate | | | |
|---|---|---|---|---|---|
| in f-roots | in hinge units | relative | fitted | observed | residual |
| -1.19 | -3.0 | .36 | 29h | 32 | +2h |
| -.86 | -1.0h | .88 | 72h | 72 | +0h |
| -.65h | 0.1h | 1.00 | 82 | 79 | -3 |
| -.48h | 1.1h | .86 | 70h | 68h | -2 |
| -.31 | 2.2 | .57 | 47 | 48h | +1h |
| -.15h | 3.1 | .34h | 28h | 27h | -1 |
| 0 | 4.0 | .16 | 13 | 12 | -1 |
| .15h | 4.9 | .06 | 5 | (7) | (2) |
| .31 | 5.8 | .01h | 1 | (1) | (0) |
| | sum | 4.25 to be 349 | 348 ✓✓ | | |
| Added to help draw the curve | | | | | |
| -1.36 | -4.0 | .16 | 13 | | |

G) SOURCES

Apparent source: K. Pearson 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen in random sampling. Philosophical Magazine 50:167- . See page 167.

Used by: F. Y. Edgeworth 191?. Article Probability in the 11th (to 14th) edition of the Encyclopedia Britannica 22:39.

M. G. Kendall and A. Stuart 1958. The Advanced Theory of Statistics. Tables 1.14 and 1.16 at pages 19 and 25 of volume 1 of 3rd edition.

T. F. Fry 192?. Probability and Its Engineering Uses.

H) PROBLEM

Repeat all calculations using pluralities, and compare: (a) quality of fit and (b) constancy of spread with what was found above.

Suspended rootogram for 4096 throws of 12 dice --
each 4, 5, or 6 counted as a success (0 to 12 possible)

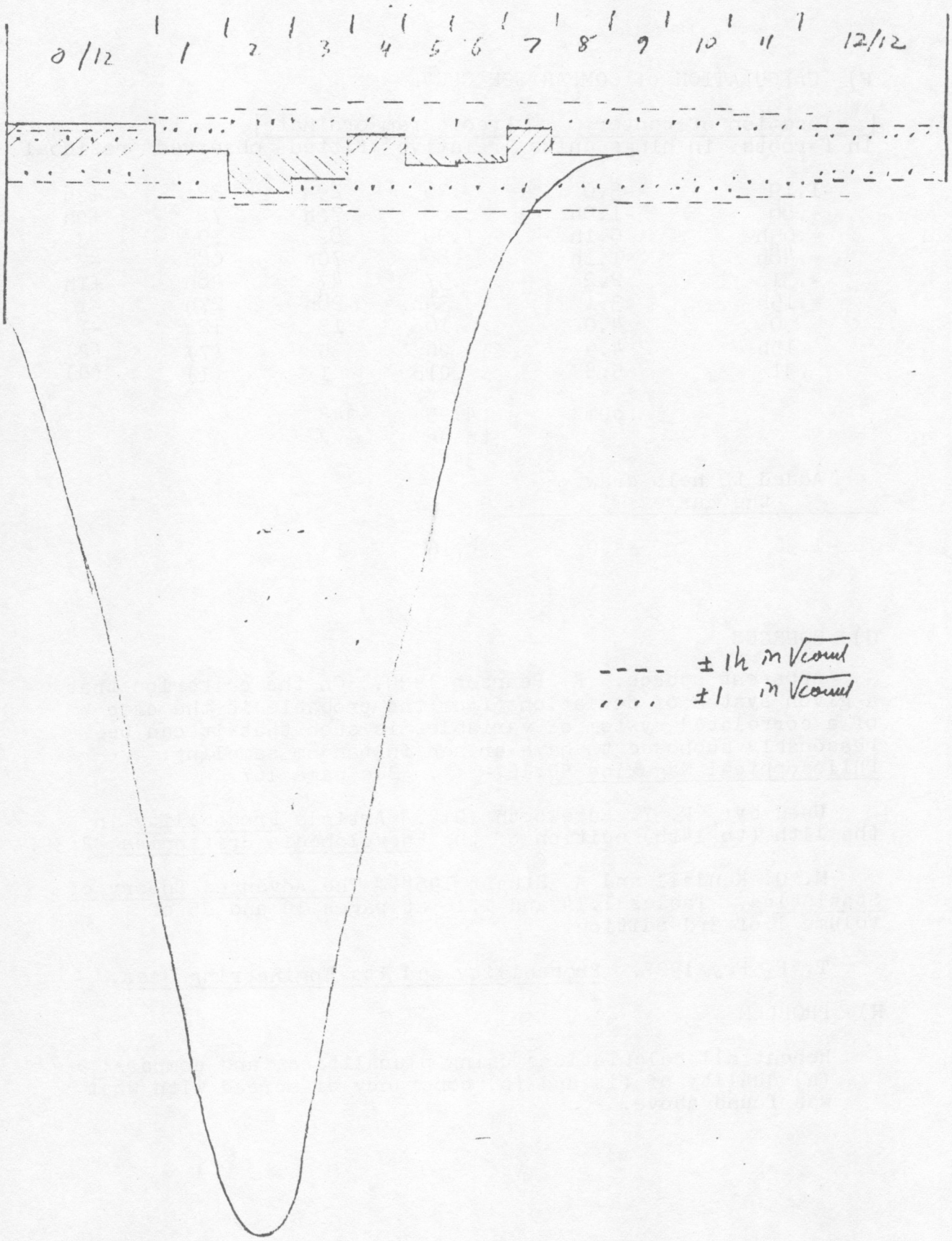
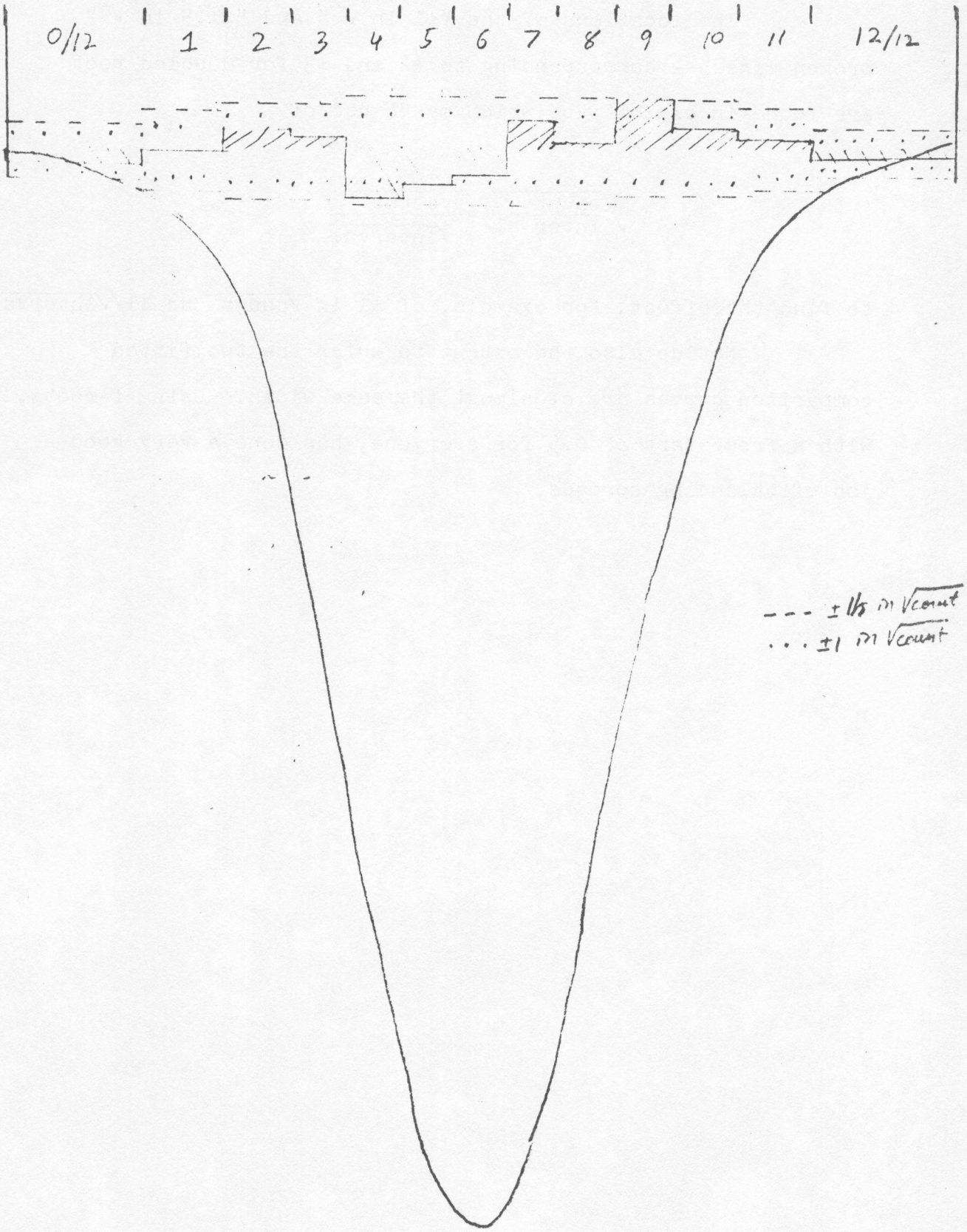


exhibit 25 of chapter 28 -- Weldon's dice

Suspended rootogram for 4096 throws of 12 dice -- each 6 counted as a success (0 to 12 possible)



given our conventions for free start -- is at least as good as we had any right to expect.

The locations of the "±1 in $\sqrt{\ }$" and "±1.5 in $\sqrt{\ }$" broken lines -- corresponding to ±2 and ±3 for doubled root are found in exhibit 26, which makes use of

$$\sqrt{\frac{\text{count}}{\text{interval}}} = \frac{\sqrt{\text{count}}}{\sqrt{\text{interval}}}$$

to find the effect, for example, of ±1 in $\sqrt{\text{count}}$ as $\pm 1/\sqrt{\text{interval}}$.

Notice also the extent to which the two fitted comparison curves are of almost the same width. Using f-roots, with a free start of 0.5 for everyone, has done a very good job of balancing spreads.

exhibit 26 of chapter 28 -- warning and control

The sizes of "±1 in $\sqrt{count}$" and "±1.5 in $\sqrt{count}$" on a plot of $\sqrt{count/interval}$ using started f-roots for 12 possible successes

| | | warning | control |
|---|---|---|---|
| successes | $\sqrt{interval}$ | $\pm 1/\sqrt{interval}$ | $\pm 1.5/\sqrt{interval}$ |
| 0 | .66 | ±1h | ±2 |
| 1 | .47 | ±2 | ±3 |
| 2 | .44 | ±2h | ±3h |
| 3 | .41 | ±2h | ±3h |
| 4 | .40 | ±2h | ±4 |
| 5 | .39 | ±2h | ±4 |
| 6 | .40 | ±2h | ±4 |
| 7 | .39 | ±2h | ±4 |
| 8 | .40 | ±2h | ±4 |
| 9 | .41 | ±2h | ±3h |
| 10 | .44 | ±2h | ±3h |
| 11 | .47 | ±2 | ±3 |
| 12 | .66 | ±1h | ±2 |

*******

28H. <u>How far have we come?</u>

This chapter has dealt with the effective expression of counted fractions.

We have been led to use either folded roots or folded logarithms for counted fractions, specifically

$$\text{f-root} = \sqrt{2f} - \sqrt{2-2f}$$

$$\text{f-log} = 1.15 \log f - 1.15 \log(1-f)$$

where f is the given fraction -- either as observed or adjusted by a free start (usually 0.25 for each class). [Table at 28-5, exhibit 1.]

Differently matched expressions are equally appropriate when equally convenient. In particular, the raw f-log

$$\log f - \log (1-f)$$

is often useful.

These expressions are sometimes useful for fractions obtained in other ways than counting.

Graphical displays using f-roots or f-logs presenting equal area for equal counts are easy. We have only to use "split cigars".

Examples where using f-roots or f-logs makes the analysis clearer and better are common.

Checking to see if a change in expression will make the analysis of a two-way table easier is not as easy for

counted fractions as it is for amounts. For elementary analyses not by means, we do well to just repeat the analysis using a different expression and then compare the results. (This comparison is best done graphically.) For elementary analyses by means we can use a short cut. (Tables at 28-26 and 28-28, exhibits 12 to 14).

We can also use f-roots and f-logs to straighten response curves.

When we need to look hard at counts of counted fractions a good thing to try is the use of started f-roots with a free start of 0.5 in each category -- 0.5 extra for those that do and 0.5 extra for those that don't.

*******

28K. Easy f-roots and f-logs with a slide rule

Tables of f-roots and f-logs are not always available. In any event they require conversion of counts into fractions or % before use. Accordingly there can be advantages to more direct calculation, especially when a slide rule is available. This is particularly so when the f-roots or f-logs to be found come in sets with the same total, something that will, in particular, occur whenever we are dealing with several alternate cutting points for a response.

Suppose, then, that, for some cutting point, our counts are a and b so that

$$f = \frac{a}{a+b}$$

$$1-f = \frac{b}{a+b}$$

$$\text{f-root} = \sqrt{2f} - \sqrt{2(1-f)}$$

$$= \sqrt{\frac{a}{(a+b)/2}} - \sqrt{\frac{b}{(a+b)/2}}$$

$$\text{or } \sqrt{2a/(a+b)} - \sqrt{2b/(a+b)}$$

Thus if a = 43, b = 471, a+b = 514, (a+b)/2 = 257 the corresponding f-root is

$$\sqrt{\frac{43}{257}} - \sqrt{\frac{471}{257}}$$

One setting of the slide rule (with many settings of the cursor) suffices to read out all the numbers needed for all calculations with the same total. (On an elementary slide rule, one resetting will be needed to overlap the other end of the scale.)

After we have set 257 on B against 1 on DF, against 43 on B we read 408 on DF and against 471 on B we read 1.353 on DF. Since

$$.408 = \sqrt{\frac{43}{257}}$$

$$1.353 = \sqrt{\frac{471}{257}}$$

we have

$$\text{f-root} = .408 - 1.353 = -.945$$

With three or more f-roots to compute with the same total, this approach should be faster than using a table to convert %s into f-roots, the %s having been found with the slide rule.

For f-logs, we write

$$\text{f-log} = 1.15\left(\log_{10}\frac{a}{a+b} - \log_{10}\frac{b}{a+b}\right)$$

$$= 1.15\left(\log_{10}a - \log_{10}b\right)$$

and calculate according to the last line. For our example we have

$$\log 43 = 1.634$$

$$\log 471 = 2.673$$

$$\text{f-log} = 1.15(1.634 - 2.673) = 1.15(-1.039) = -1.195$$

(The multiplications by 1.15 -- if not omitted to give a raw f-log -- may either all be postponed to the end of the calculation -- and then be done together -- or (if D and L scales are both available on the body of rule) the rule can be left set for multiplication by 1.15.)

This procedure, too, is likely to be easier than using a table.

*****

28L. Comparison of simple and refined expressions for counted fractions (background)

We have confined ourselves to three very simple ways of expressing fractions:

- pluralities - folded fractions
- f-roots = folded square roots
- f-logs = folded logarithms.

We don't have to stop here -- we can form and fold any fractional power we choose. If we want to match behavior near 50% we should use

$$\frac{2^\lambda}{2\lambda}\left(f^\lambda - (1-f)^\lambda\right)$$

which is often most conveniently computed as

$$\frac{(2f)^\lambda - (2-2f)^\lambda}{2\lambda}$$

for whatever $\lambda$ we may choose. We will call what we get this way "folded $\lambda$-powers". Thus f-roots are folded 0.5-powers and f-logs play the role of folded zero-powers. (Recall from chapter 3 that -- as a function, properly scaled -- the log fills in neatly between small positive powers and small negative powers.) Folded powers give us a large and convenient family of ways of expressing fractions.

Three more complicated -- or perhaps refined -- ways of expressing fractions are frequently used -- or have been recently suggested. Some of the reasons for their choice lie outside our scope, but we can surely look at their values, comparing them with one another and with appropriate folded fractional powers.

One is known under such names as "the angular transformation", "anglits", "sinits", or just "angles". Its mathematical expression, when matched near 50% is

$$\theta = 2(\text{the angle whose sine is } \sqrt{f}) - (\text{a right angle})$$

with all angles measured in radians (1 radian = 57.3 degrees).

A second way of expression is known under such names as "probits" and "normits", and is associated with the "inverse of the cumulative Gaussian distribution". (If you know what this is well and good -- if you do not, don't worry.) This choice is very widely used, but has no really simple mathematical formula. The simplest we can do, if we want to match near 50% is to use:

$$\text{matched normit} = \sqrt{\frac{2}{\pi}} \cdot x$$

where

$$\text{fraction} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

(Again if you recognize these expressions, fine -- if you do not, relax and go on.)

A third refined way of expression has recently been suggested by Cox and Snell. It can be expressed in terms of an incomplete beta function of fractional order. (What can we helpfully say further?) Again we are free to look at its numerical values and see how it behaves.

Finally we note that a fair amount of use is given to "logits", which can be defined by

$$\text{fraction} = \frac{1}{1 + e^{-(\text{logit})}}$$

This is, however, nothing new to us, since

$$(\text{logit}) \equiv 2\cdot(\text{f-log})$$

where the 1.15 factor is included in the f-log, is an identity.

Exhibit 27 compares pluralities, f-roots, f-logs, three carefully selected fractional powers, the matched forms of anglits, normits and logits, and the Cox-Snell suggestion. We see that each of the more refined ways of expressing a fraction can be quite closely matched by an appropriate folded power.

Looking in more detail, we see that: The Cox and Snell suggestion is matched by the folded power with $\lambda = .57$ to within ±0.01 over the entire range. Anglits are matched by the folded power with $\lambda = .41$ to within ±0.01 from 0.1% to 99.9%, and to within ±0.05 overall. Normits are matched by the folded power with $x = .14$ to within ±0.01 from 0.5% to 99.5%, and within ±0.03 from 0.1% to 99.9%. (This degree of matching is all the more striking when we note that these two have to differ by an infinitely large amount for the most extreme fractions -- 0% or 100%.)

Exhibit 27 of chapter 28

Summary comparison of ways of expressing fractions

| + | Plur. | λ = .57 | (1) | f-root | λ = .41 | (2) | λ = .14 | (3) | f-log | - |
|---|---|---|---|---|---|---|---|---|---|---|
| 50% | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | 50% |
| 60% | .20 | .20 | .20 | .20 | .20 | .20 | .20 | .20 | .20 | 40 |
| 70% | .40 | .41 | .41 | .41 | .41 | .41 | .42 | .42 | .42 | 30 |
| 80% | .60 | .63 | .63 | .63 | .64 | .64 | .67 | .67 | .69 | 20 |
| 85% | .70 | .75 | .74 | .77 | .77 | .78 | .83 | .83 | .86 | 15% |
| 90% | .80 | .88 | .88 | .89 | .92 | .93 | 1.03 | 1.02 | 1.10 | 10 |
| 92.5% | .85 | .96 | .95 | .97 | 1.01 | 1.02 | 1.15 | 1.15 | 1.26 | 7.5 |
| 95% | .90 | 1.03 | 1.03 | 1.06 | 1.11 | 1.12 | 1.32 | 1.31 | 1.47 | 5 |
| 96% | .92 | 1.07 | 1.07 | 1.10 | 1.16 | 1.17 | 1.41 | 1.40 | 1.59 | 4% |
| 97% | .94 | 1.10 | 1.11 | 1.15 | 1.21 | 1.22 | 1.51 | 1.50 | 1.74 | 3 |
| 98% | .96 | 1.14 | 1.15 | 1.20 | 1.28 | 1.29 | 1.65 | 1.64 | 1.95 | 2 |
| 99% | .98 | 1.20 | 1.20 | 1.27 | 1.37 | 1.37 | 1.87 | 1.86 | 2.30 | 1 |
| 99.5% | .99 | 1.23 | 1.24 | 1.31 | 1.43 | 1.43 | 2.06 | 2.06 | 2.65 | 0.5% |
| 99.9% | .98 | 1.26 | 1.27 | 1.37 | 1.52 | 1.51 | 2.44 | 2.47 | 3.45 | 0.1 |
| 100.0% | 1.00 | 1.30 | 1.29 | 1.41 | 1.62 | 1.57 | 3.57 | ∞ | ∞ | 0.0 |

(1) This is Cox and Snell's suggestion, matched near 50%.

(2) These are anglits (which are already matched).

(3) These are matched probits or normits.

Source for (1): D. R. Cox and E. J. Snell 1968. A general definition of residuals. J. Royal Statis. Soc. Ser. B. 30:248-265. Table 2 at page 260.

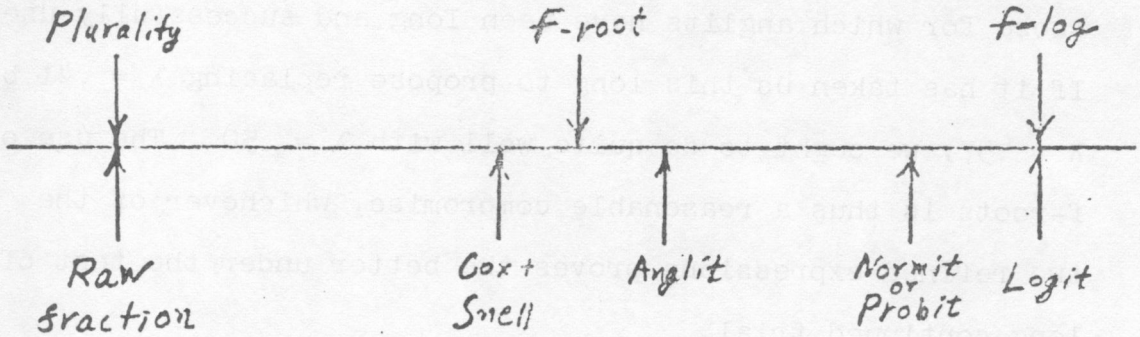
As a result, we may take the matching values of $\lambda$ as good indicators of how the more refined expressions behave. Exhibit 28 pictures the position, and suggests why f-root and f-log seem likely to meet almost all our needs.

The Cox and Snell suggestion is new -- as yet untested by extensive experience. Its main intended uses are those for which anglits have been long and successfully used. If it has taken us this long to propose replacing $\lambda = .41$ by $\lambda = .57$, we ought to do quite well with $\lambda = .50$. The use of f-roots is thus a reasonable compromise, whichever of the two refined expressions proves the better under the test of long-continued trial.

Of course, f-logs match logits exactly. They are reasonably close to normits-probits. There may well be situations where the latter will be appreciably better than either f-logs or f-roots, but we have every reason to think that these situations will be uncommon. In most situations we expect to do quite well with one of the three simple expressions -- pluralities, f-roots or f-logs. (If we need to use some other expression, normits are a reasonable candidate, but so are folded powers for a wide variety of values of $\lambda$.)

exhibit 28 of chapter 28 -- fractions re-expressed

Picture of the relationship among the simple and refined expressions for counts



*****

28M. Fitting comparison curves related to folded powers (optional)

Every form of expressing fractions corresponds to a shape of distribution. For we can start with a rectangular distribution extending from 0 to 1, and say:

- if the raw fraction has a rectangular (uniform) distribution, what then is the distribution of our chosen expression?

Doing this with folded powers - including folded logs as a special case, produces an interesting and useful set of distributions. Those that are most easily manageable are those corresponding to powers, 1, 1/2, 0, -1/2, and -1. Fortunately these cover a quite wide range of symmetrical shapes, and come close enough to one another so that each of a wide variety of purposes can be served by some one of them.

Just as each of these expression stretches out the extremes more than the preceding one, so now each of these new distributions will be more stretched out -- and their comparison curves lower toward their ends -- than the preceding ones.

Distributions of the shape arising for $\lambda = 0$ have been used for a long time under the name "logistic". By analogy, then, we may use "rootistic" for $\lambda = 1/2$, "tooristic" for $\lambda = -1/2$ and "invistic" for $\lambda = -1$, where we have turned "root" around as "toor" to indicate the change from a positive to a negative power. The shapes of these distributions

are shown in exhibits 29 and 30, where the hinge positions and central heights have been matched.

As $\lambda$ decreases, we can see the tails growing longer and longer -- and longer. With the hinges in fixed position, the points cutting off 10% on each tail, shown by the change from single to double shading, move further and further out as $\lambda$ moves from +1 through 0 to -1.

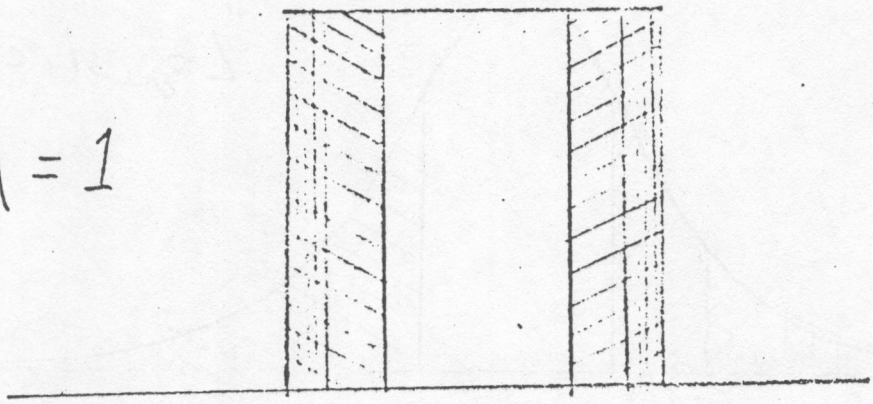
It should not surprise us that this connection between distributions, on the one hand, and expressions of fractions, on the other, is a two-way street. Not only can we get distributions from expressions of fractions -- we can also -- as people have long done -- get expressions for fractions from distributions. It is from the Gaussian distribution that normits and probits -- just mentioned in section 28L as expressions for fractions -- were gotten.

Since normits and probits can be closely approximated by folded powers with $\lambda = 0.14$, it follows that the distribution of the sort we are here considering for $\lambda = 0.14$ is a close approximation to the Gaussian distribution. It happens to be true that the distribution for $\lambda = -0.97$, which is close to the invistic distribution, is a good approximation to the Cauchy distribution, which we took in section 26C as a fine example of a long-tailed distribution. Thus the family of $\lambda$-distributions, with selected instances from which we have just been concerned, comes close to the Gaussian

exhibit 29 of chapter 28 -- comparative delays

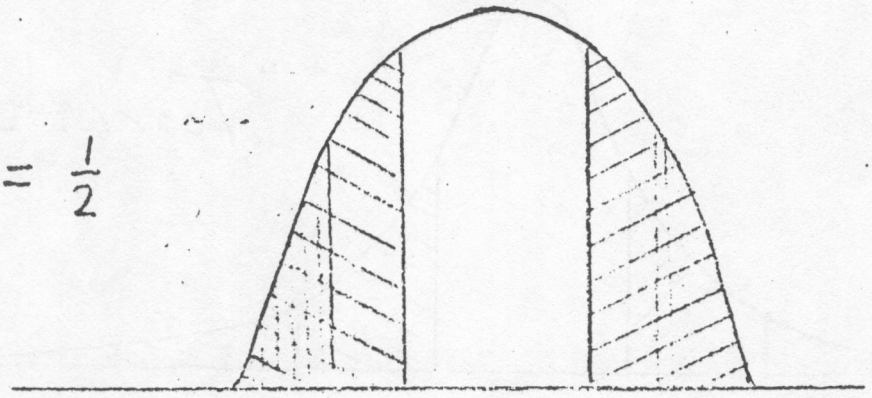
Some distributions related to folded powers

$\lambda = 1$



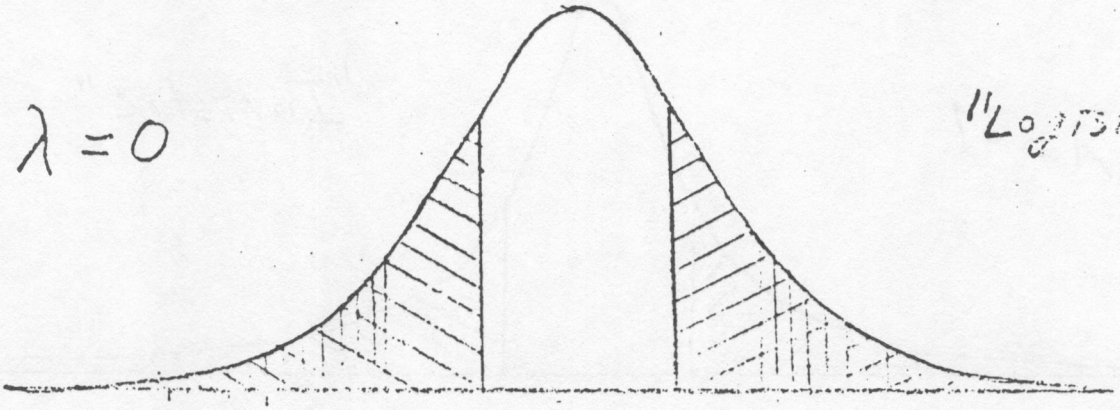
"Rectangular"

$\lambda = \frac{1}{2}$



"Roofstric"

$\lambda = 0$

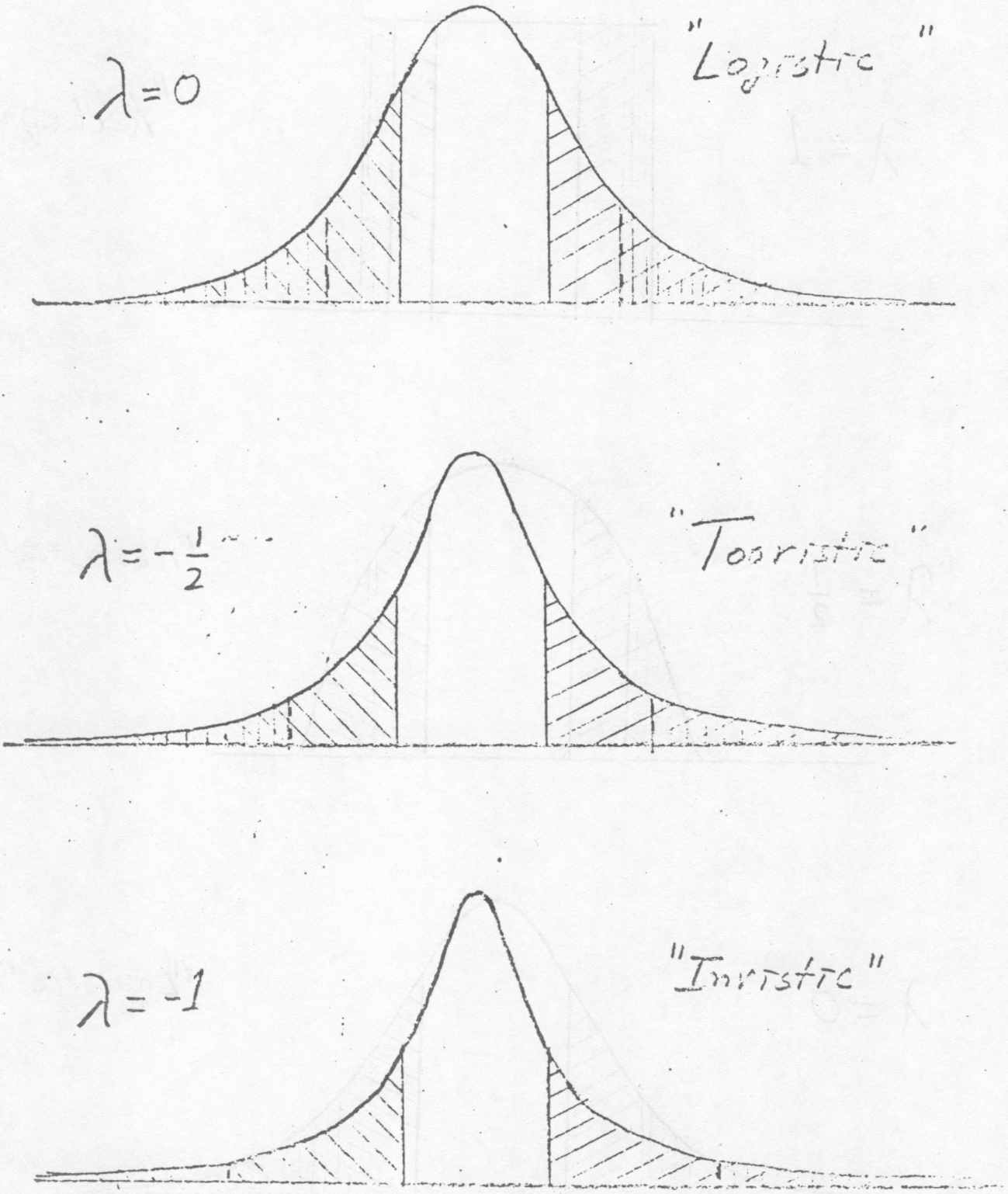


"Logistic"

Note: Middle 50% unshaded, both end 10%'s dark shaded.

exhibit 30 of chapter 28 -- comparative displays

More distributions related to folded powers




Note: Middle 50% unshaded, both end 25's doubly shaded.

and the Cauchy distributions, as well as including the rectangular.

Exhibit 31 summarizes the behavior of the λ-distributions for λ = 0.5, 0, -0.5 and -1 on a scale based on the half mid-spread, both as to height of curve and as to remaining fractionsof area.

For those who need more detail (in particular for the fitting of comparison curves in the style of 26C), exhibit 32 gives more detailed numerical values.

Summary behavior of λ-distributions

| | Relative height of curve | | | | | | Fraction of area beyond x | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| x | λ=1 | 0.5 | λ=0 | -0.5 | λ=-1 | x | λ=1 | 0.5 | λ=0 | -0.5 | λ=-1 | x |
| 0.0 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.0 | .500 | .500 | .500 | .500 | .500 | 0.0 |
| 0.5 | 1.00 | .97 | .93 | .85 | .75 | 0.5 | .375 | .372 | .366 | .352 | .351 | 0.5 |
| 1.0 | 1.00 | .90 | .75 | .59 | .45 | 1.0 | .250 | .250 | .250 | .250 | .250 | 1.0 |
| 1.5 | 1.00 | .76 | .54 | .39 | .27 | 1.5 | .125 | .142 | .161 | .178 | .19 | 1.5 |
| 2.0 | cuts | .54 | .36 | .25 | .18 | 2.0 | .000 | .057 | .110 | .131 | .153 | 2.0 |
| 2.5 | .00 | .21 | .23 | .17 | .13 | 2.5 | " | .007 | .060 | .100 | .128 | 2.5 |
| 3. | | .00 | .14 | .12 | .09 | 3. | | none | .036 | .078 | .110 | 3. |
| 4. | | " | .05 | .07 | .06 | 4. | | " | .012 | .051 | .086 | 4. |
| 5. | | " | .016 | .04 | .04 | 5. | | " | .004 | .036 | .069 | 5. |
| 6. | | " | .006 | .02 | .03 | 6. | | " | .001 | .027 | .059 | 6. |
| 7. | | " | .002 | .017 | .02 | 7. | | " | .000 | .021 | .051 | 7. |
| 8. | | " | .001 | .012 | .016 | 8. | | " | " | .017 | .045 | 8. |
| 10. | | | .000 | .006 | .010 | 10. | | | | .011 | .036 | 10. |
| 15. | | | " | - | .005 | 15. | | | | - | .025 | 15. |
| 20. | | | " | - | .002 | 20. | | | | - | .019 | 20. |

Notes: (1) x is in units of the half midspread.
(2) some values for λ = -1.0 are interpolated.

exhibit 32 of chapter 28 -- lambda behavior

Information on λ-distributions suitable for fitting comparison curves (Entries are relative heights, followed in parentheses by remaining fractions)

| x | λ=0.5 | λ=0 | λ=-0.5 | λ=-1 | x |
|---|---|---|---|---|---|
| 0.0 | 1.00 (.500) | 1.00 (.500) | 1.00 (.500) | 1.00 (.500) | 0.0 |
| 0.1 | 1.00 (.476) | 1.00 (.473) | .99 (.470) | - | 0.1 |
| 0.2 | .99 (.448) | .99 (.445) | .97 (.441) | .95 (.434) | 0.2 |
| 0.3 | .99 (.423) | .97 (.418) | .94 (.413) | - | 0.3 |
| 0.4 | .98 (.397) | .95 (.392) | .90 (.384) | .83 (.375) | 0.4 |
| 0.5 | .97 (.372) | .93 (.366) | .85 (.358) | - | 0.5 |
| 0.6 | .96 (.347) | .90 (.341) | .80 (.333) | .68 (.325) | 0.6 |
| 0.7 | .95 (.322) | .87 (.317) | .75 (.310 | - | 0.7 |
| 0.8 | .93 (.297) | .83 (.293) | .70 (.289) | .55 (.283) | 0.8 |
| 0.9 | .92 (.273) | .79 (.271) | .64 (.268) | - | 0.9 |
| 1.0 | .90 (.250) | .75 (.250) | .59 (.250) | .45 (.250) | 1.0 |
| 1.1 | .87 (.227) | .71 (.230) | .55 (.233) | - | 1.1 |
| 1.2 | .85 (.205) | .67 (.211) | .50 (.217) | .37 (.223) | 1.2 |
| 1.3 | .82 (.183) | .62 (.293) | .46 (.203) | - | 1.3 |
| 1.4 | .79 (.162) | .58 (.177) | .42 (.190) | .30 (.201) | 1.4 |
| 1.5 | .76 (.132) | .54 (.161) | .39 (.178) | - | 1.5 |
| 1.6 | .72 (.123) | .50 (.147) | .35 (.167) | .25 (.182) | 1.6 |
| 1.7 | .68 (.105) | .46 (.134) | .32 (.157) | - | 1.7 |
| 1.8 | .64 (.088) | .43 (.122) | .30 (.148) | .21 (1.67) | 1.8 |
| 1.9 | .59 (.072) | .39 (.110) | .27 (.139) | - | 1.9 |
| 2.0 | .54 (.057 | .36 (.100) | .25 (.131) | .18 (.154) | 2.0 |
| 2.1 | .49 (.044) | .33 (.091) | .23 (.124) | - | 2.1 |
| 2.2 | .43 (.032) | .30 (.082) | .22 (.117) | .16 (.142) | 2.2 |
| 2.3 | .36 (.022) | .27 (.074) | .20 (.118) | - | 2.3 |
| 2.4 | .29 (.013) | .25 (.067) | .19 (.105) | .14 (.132) | 2.4 |
| 2.5 | .21 (.007) | .23 (.060) | .17 (.100) | - | 2.5 |
| 2.6 | .13 (.002) | .20 (.054) | .16 (.095) | .12 (.124) | 2.6 |
| 2.7 | .03 (.000) | .19 (.049) | .15 (.090) | - | 2.7 |
| 2.8 | .00 (.000) | .17 (.045) | .14 (.086) | .11 (.116) | 2.8 |
| 2.9 | " | .15 (.040) | .13 (.082) | - | 2.9 |
| 3.0 | " | .14 (.036) | .12 (.078) | .09 (.110) | 3.0 |

(Cont'd)

exhibit 32 (Cont'd.) - 2

| x | λ=0 | λ=-0.5 | λ=-1 | |
|---|---|---|---|---|
| 3.0 | .137 (.036) | .120 (.078) | .094 (.110) | 3.0 |
| 3.2 | .112 (.029) | .106 (.071) | .085 (.104) | 3.2 |
| 3.4 | .090 (.023) | .093 (.065) | .076 (.098) | 3.4 |
| 3.6 | .073 (.0188) | .082 (.060) | .069 (.093) | 3.6 |
| 3.8 | .060 (.0151) | .073 (.056) | .062 (.089) | 3.8 |
| 4.0 | .048 (.0112) | .065 (.051) | .057 (.085) | 4.0 |
| 4.2 | .039 (.0110) | .058 (.048) | .053 (.081) | 4.2 |
| 4.4 | .031 (.0079) | .052 (.044) | .048 (.078) | 4.4 |
| 4.6 | .025 (.0063) | .047 (.041) | .046 (.075) | 4.6 |
| 4.8 | .020 (.0051) | .043 (.039) | .041 (.072) | 4.8 |
| 5.0 | .016 (.0041) | .039 (.036) | .038 (.069) | 5.0 |
| 5.5 | .009 (.0024) | .031 (.031) | - | 5.5 |
| 6.0 | .006 (.0014) | .025 (.027) | .027 (.058) | 6.0 |
| 6.5 | .003 (.0008) | .020 (.024) | - | 6.5 |
| 7.0 | .001 (.0005) | .017 (.021) | - | 7.0 |
| 7.5 | .001 (.0003) | .014 (.019) | - | 7.5 |
| 8.0 | .000 (.0002) | .012 (.017) | .016 (.045) | 8.0 |
| 8.5 | " (.0001) | .010 (.015) | - | 8.5 |
| 9.0 | " (.0001) | .009 (.013) | - | 9.0 |
| 9.5 | " (.0000) | .007 (.012) | - | 9.5 |
| 10.0 | " (.0000) | .006 (.011) | .010 (.036) | 10.0 |

λ = 0 is "logistic".

λ = -0.5 is "tooristic".

λ = -1 is "invistic".

*******

28P. Problems for chapter 28

1. Do the problem of exhibit 23, page 28-55.

2. Take the 8×3 table in panel A of exhibit 15 and analyze by middle mean differences. Find the 9×2 table corresponding to panel B of exhibit 16, and the change column.

3. Watson reports the results of an experiment in which aphids were fed for one period on an infected plant and then for a second period on a healthy plant, the response being the number of healthy plants, out of 30, later found to be infected. The data are in exhibit 33. What form of expression seems indicated? To what appearances does your analysis lead you?

exhibit 33 of chapter 28 -- aphis infecting

Watson's data on number of plants out of 30 infected with plant virus hy. III after aphis feeding

| Time on infected plant | Time on healthy plant | | | | | |
|---|---|---|---|---|---|---|
| | 2 min | 5 min | 15 min | 1 hr | 6 hr | 12 hr |
| 2 min | 17 | 20 | 13 | 18 | 16 | 23 |
| 5 min | 11 | 13 | 10 | 14 | 18 | 15 |
| 15 min | 10 | 6 | 10 | 9 | 9 | 18 |
| 1 hr | 3 | 3 | 2 | 2 | 7 | 4 |
| 6 hr | 3 | 1 | 7 | 6 | 6 | 12 |
| 12 hr | 6 | 9 | 8 | 2 | 4 | 9 |

Source: M. A. Watson 1936. Factors affecting the amount of infection obtained by Aphis transmission of the virus Hy. III. Phil. Trans. Roy. Soc. London 226: 457-489, expecially table XI at page 477.